

220 AGQPVVLNPNRLVSAHYATSVPTWFSDFPTPNM-PCINKNQKGYLNFQNIAPV 270
 441 VADEFWISAPDKLVYSVHYDGPDIYMQPFKKDQDINTLYEECVPNWYIIVEQNIAPM 500
 279 WLGEFGTTLQSTDTQTLKTLVQYLRPTAQYAGDSFQWTFWNNPDSGDTGILKDDWOT 338
 501 LIGEGGKLIENNRKWLCELATFI-----AEKLLHHTFWAFNPNSADTGGLMLEDWKT 554
 339 VDTWKDGYLAP 349
 555 VDEEKYAIIVP 565
 RESULT 14
 Q9EYQO PRELIMINARY; PRT; 534 AA.
 Q9EYQO:
 01-WAR-2001 (TrEMBLrel. 16, Created)
 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 Cellulase Cel5-N.
 CELN.
 Clostridium cellulolyticum.
 OS
 Clostridia; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 Bacteria; Clostridium.
 OC
 NCBI_Taxid=1521;
 [1]
 SEQUENCE FROM N.A.
 RNP Belaich A., Gaudin C., Gal L.;
 RA "A gene cluster of cellulases of Clostridium cellulolyticum";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DE EMBL: AF316823; AAG45162.1; --
 DR HSSP: P54593; LECE.
 DR GO: GO:0005509; F:calcium ion binding; IEA.
 DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO: GO:0000272; P:polysaccharide catabolism; IEA.
 DR InterPro: IPR002105; Dockerin_1.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001547; Glyco_hydro_5.
 DR Pfam: PF00150; cellulase; 1
 DR Pfam: PF00404; Dockerin_1; 2.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR SEQUENCE 534 AA; 59671 MW; 980785F0971E5A59 CRC64;
 Query Match 30.3%; Score 602.5; DB 2; Length 534;
 Best Local Similarity 33.0%; Pred. No. 1.7e-39;
 Matches 134; Conservative 60; Mismatches 143; Indels 69; Gaps 13
 QY 6 W-HTSGREILDANNVPVRIAGINWGFPTCNVYVHGLWSRDYRSMLDQIKSLGNTIRLP 64
 Db 38 WLHCVGDKTIYDNGREVMLTGANWGFNCSENVFHGA- YDVKNILTSVADRGIGLLRVP 96
 QY 65 YGDDIL-----KPG-----TVPNSINFRQNNQDLOG-----LTSQVMDKIVAYAGQI 107
 Db 97 ISTELLYSMWTGKPNKVSSTASNNPPYTVVNPFDYDPATDGPKNSEMEIFDIIMKYCKEL 156
 QY 108 GLRIILDRHPPDCSQS---ALWY-----TSSVSEATWISDLQALAQRYKGNPTVVGF 158
 Db 157 GIKWMTDVHSPDANNSGHYPWLYGLETTAGMITTDKIDITLTWLAGKYNDITLAI 216
 QY 159 LKNEPH-----DPACWGCGDPSIDRLAAERAGNAVLNVSNPLLIFVEGVQSY- 206
 Db 217 LKNEPHGKGYTNAAPTDWAKWDNTDENNKYAAERCSKEILAVNPKLLIMIEGBOYP 276
 QY 207 -----NGDSY-----WGGNLOGAGQYPVVLNYPN-RLVYSAEDYATSV 244
 Db 277 KTEKGYFTFPDVGASGDAAFWHGSGWGNLGRVKQYPIDLGPLNSQIVTSPHYGFSV 336
 QY 245 YPOTWF-SDPTFPNNFEG;WNKNWGYLFNQNIAPVWLGFEFGTTLQSTDTQTLKTLVQYL 303
 Db 337 YNCSWFKPATTCTIADDYVYDTWYVIDDKTAPLLIGWGGFDGAKNSKMTLLRDYM 396

```

QY      304 RTAQYGADSFQWTFWSNWPDSGDTGGILKDDMQTVDTVKDGYLAP 349
DB      397 I-----KNRINHFWCLPNPNSGDTGLIGNDMSWTDEEKYGLKP 436

RESULT 15
Q7X2N2
ID       Q7X2N2 PRELIMINARY;          PRT;    616 AA.
AC       Q7X2N2;
DT       01-OCT-2003 (TriEMBLrel. 25, Created)
DT       01-OCT-2003 (TriEMBLrel. 25, Last sequence update)
DT       01-OCT-2003 (TriEMBLrel. 25, Last annotation update)
DB       Endoglucanase.
GN       GN
GC       CEL5B.
OS       Thermomonospora fusca.
OC       Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales
OC       Streptoporangineae; Nocardiopsaceae; Thermobifida.
OX       NCBI_TaxID=2021;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN=TW51;
RA       Posta K., Beki E., Kukolyva J., Hornok L.;
RT       "Phylogenetic relationships of Tf cel5B, a new endoglucanas
RT       gene from thermobifida fusca.";
RL       Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR       EMBL; AY298814; AAP56348.1; -.
SQ       SEQUENCE   616 AA;  67701 MW;  24FFC1EAlA13F5639 CRC64;

Query Match           29.4%; Score 583; DB 2; Length 616;
Best Local Similarity 34.08; Pred. No. 7.3e-38;
Matches 129; Conservative 63; Mismatches 149; Indels

QY      1 AGGGYW-HTSGREILDANNVPVRIAGINWGFETCNYYVHGLWSRDYSRM
DB      33 SGTADMHLTDGNRIVDSAGNEVWLITGANWFGNTSERMFHGLWAANIEDII
QY      60 TIRLPYSDDL---KPGTM-PNSINFRMNODLQLGTSLOVMDKIVAYAGA
DB      93 MYRVPISTQLLEWKNGQGSPGVN-EYNPELAGMTLEVDFDYWLQJLCE
QY      116 H--RPDCSGO-SALWYTSSVSEATWISDLQAARYKGNPVTVGFDPLHNES
DB      152 HSAEADNSGHYPVYWKGDITTFDYTAWEVWTERYKKNDDTIVAADIKNKE
QY      167 ACWCGGDPSTD-WELAAERAGNAVLSVNPILLIFVEGVQSXNGD-----
DB      212 AKWD-GSTDINDFKHVCEIAGKRIILAINPNMLILCEGLEIYPDGQGDWSSS
QY      213 WGNLQAGQGPVVVLNV-FNELVISAHDYATSVYPTWFSDPTFPNNM-F
DB      271 WGNLRGVADHPVDLGAHQQDLVYSPHDYGPSVFEPQPFEGEWNRQTLE
QY      271 FNONIAPVLLGEFGTTLOSTDQTKLVQLVLRFTAQYGADSFQWTFWSN
DB      331 HEDDIAPFLLEGWGFGDLGGDNKWMNTALRSIID-----EKMHHTFWFA
QY      331 ILKDDWQTVDTVKGYLAP 349
DB      385 LLNDWTWTDEAKYAFLXP 403

Search completed: August 2, 2004, 16:36:58
Job time : 37.3333 secs

```

```

Db 829 RGVKYPINLKYQNKVYSPHDYQPSYVQOPWFY-PQFTKESILLQDCWRPNWAYIMEEN 887
Qy 275 IAPVWLGFYFTLSTDTQTLVQLVLRPTAQYAGDSFQWTFWNSWNPDSGDTGGILKD 334
Db 888 IAPLLIGWGGHLCADNEKMKYLRDYII-----ENHIHTWCFNANSNGDTGGILGY 941
Qy 335 DWQTVDTVKDGYLAP 349
Db 942 DFTTWDEKKYSLKP 956

RESULT 12
ID Q9X3P6 PRELIMINARY; PRT; 1426 AA.
AC Q9X3P6;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE CelB.
GN CelB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D.; Reeves R.A.; Farrington G.K.; Anderson P.; Williams D.P.;
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340 (2000).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
CC EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR PRINTS; PR00134; GLYDRLASE10.
DR PRODOM; PD001947; CBD_3; 3.
DR SMART; SMO0633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 31.4%; Score 623; DB 2; Length 1426;
Best Local Similarity 35.5%; Pred. No. 1.5e-40;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

Qy 10 GREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLPYSDDI 69
Db 1022 GNKIVDKGKRVLTGIVNGFNTGTNTVFCGWSCNLKSALAEIARGFNLVRPISAEI 1081

Qy 70 L---KPGTMPN-SINFRQMNQDGLTSLQWMDKIVAYAGQIGLRILDRH--RPDCSG 123
Db 1082 ILNWSKGYPKPNINY-VYNPELEGLTSLVDFVVKTCKEVGKIMLDIHSKTDAMGH 1140

Qy 124 -SALWYTSYSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCDDPS 175
Db 1141 IYPWYTDITPDIYKACEWITERYKNDTITVAFDLKNPEHKGPKQDVSFAKWDNSTDI 1200

```

```

Qy 176 IDWELAAERAGNAVLVSNPNLLIFVEGVQSYNGD-----SYWNGNQLQAGQ 222
Db 1201 NNWKYAARECAKRIILAKNPNMLIVIEGIEAYPKDDVVTWTSKSSDYYSYTWGNGNLRGVKK 1260
Qy 223 YPVVL-NVFNRLVYSAHDYATSVYPTW----FSDPTFNNMFGIWNKMGVLFNQNIAP 277
Db 1261 YPINLQYQNKVYSPHDYQPSYVQOPWFY-PQFTKDYLYND---CWRDNWTYIMDNGIAP 1317
Qy 278 VWLGEFGTTLQSTDTQTLVQLVLRPTAQYAGDSFQWTFWNSWNPDSGDTGGILKDDWQ 337
Db 1318 LLIGEWGGYLDGDNKMKWYLRDYII-----ENHIHTWCFNANSNGDTGGILGYDFS 1371
Qy 338 TVDTVKDGYLAP 349
Db 1372 TWDEQKYNFLKP 1383

RESULT 13
Q9L3J2
ID Q9L3J2 PRELIMINARY; PRT; 660 AA.
AC Q9L3J2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE 1,4-beta-cellobiohydrolase.
GN CELO.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E7;
RC Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RL EMBL; AJ275975; CAB76938.1; -.
DR HSSP; P54583; IECB.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR000437; Prok_lipoprot_5.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hydrolase.
SQ SEQUENCE 660 AA; 75247 MW; 04A042002C288CFC CRC64;

Query Match 31.3%; Score 621.5; DB 2; Length 660;
Best Local Similarity 35.8%; Pred. No. 6.9e-41;
Matches 133; Conservative 64; Mismatches 141; Indels 33; Gaps 9;

Qy 6 W-HTSGRIILDANNVPVRIAGINWFGFETCN-VYVHGLWSRDYRSLMDQIKSLGYNTIRL 63
Db 201 WLHVEGNLIKDAQNGTVYLTGIVNGFETDANGPHGLKNCNLEDSLDLMAKLGFNILRI 260
Qy 64 PYSDDIL--KPG-TMPNSINFRQMNQDGLTSLQWMDKIVAYAGQIGLRILDRH-RP 118
Db 261 PISAEIILQWKNKGERVEISFVNTYENPRDLGLSSLEILDYINEMKNGKMGAMIDMHSS 320
Qy 119 DCSGQSALWYTSYSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----DPACWCG 172
Db 321 KDSYGENLWYNKNDITMEEFIEAWKIWERYKDDDTVIADVLDKNEPHGKYSGNIAKMDSD 380
Qy 173 DPSIDWRLLAAERAGNAVLVSNPNLLIFVEGVQSYNGDSY-----WVGNGNLQ 219
Db 381 NDPNNWKKAAEIIAEIILAINPILLIVVEGEAYPMEGYDYNCGEFTTYCNWNGNLRG 440

```

```
QY 338 TUDTKDGLAP 349
DB 377 TIDTKLALVOP 388

RESULT 10
Q87AH4 PRELIMINARY; PRT; 614 AA.
ID Q87AH4
AC Q87AH4;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN XYLCA OR PD1851.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.B.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco B.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
DR EMBL; AB012560; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001319; Bac celose-metab.
DR InterPro; IPR008965; Cellulbind.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRINTS; PR01228; EGGSHLL.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 62041 MW; 29EC67E42F4BC3F7 CRC64;

Query Match 33.1%; Score 656.5; DB 16; Length 614;
Best Local Similarity 39.9%; Pred. No. 1e-43;
Matches 141; Conservative 60; Mismatches 123; Indels 29; Gaps 12;

QY 5 YWHTSGRIIDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLODKSLGYNIRLP 64
DB 26 YSISHGKVVDDXGN-QIGLRGVNWFGETGTHVNVGLWLNWKEFIFQLCGMGNIRLP 84
QY 65 YSDDLKPGTWPNSINFRQNDLQGLTSLOWDKIVAYAGQIGLRILDRHPCDCGOS 124
DB 85 FCPANLNGTSPSSIDYSR-NPDLQGLSSLQILDKVVKELSDRMVYLLDHRHPCDCSAIS 143
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184
DB 144 ELWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 203
QY 185 AGNAVLSVNPILLIFVEGVQSYNGD-----YWGNGNLQAGQYPPVLYNP-NRLY 235
DB 204 AAAALEAAPKWLIVGEGI-----GENPICSTIGHFWENLEPMDCTP--LKVPADHLL 257

QY 236 SAHDYATSVYPQTWFSPTFPNNMGIWNKNGYLFNQNIAPVWLGEFGTTL--QSTTDQ 293
DB 258 MPHYYGPDVYQYPFNSPDPFNNMAALWDKHFHAKAGYA-NAIGFSGYKGGDPDI 316
QY 294 TALKLYQLAPTAQYG-ADSFQWTFNSWPDSDGTGGLIKDDQIV--DTVK 343
DB 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSVDTCGMVGNWDWTFPRDDKVK 363

RESULT 11
O24820 PRELIMINARY; PRT; 1000 AA.
ID O24820
AC O24820;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
RT thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AB080029; BAA22939.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actinin.
DR InterPro; IPR001956; CBM 3; 1.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 31.7%; Score 629.5; DB 2; Length 1000;
Best Local Similarity 35.5%; Pred. No. 2.8e-41;
Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HTSGRIIDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLODKSLGYNIRLP 64
DB 590 WLYVSGNKIVDXKGRPVWLAGVNWFGYNTGNVFDGVWSCNLSKTLAEIANRGNLRVP 649
QY 65 YSDDL---KPGTWP-N-SINFRQNDLQGLTSLOWDKIVAYAGQIGLRILDRH--RP 118
DB 650 ISAILNWSGIIYKPFNIN-YNPELGKNSLEVPDIVVQTCVKGLKIMLDHSIKT 708
QY 119 DCSQ-SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWG 170
DB 709 DAMGHYPPWYDKFTDFYKACEWITNKNDDIIIAFDLKNEPHGRPWQDTTFAKWD 768
QY 171 CGDSIDWRLAABRAGNAVLSVNPILLIFVEGVQSYNGD-----SYWGGNLL 217
DB 769 NSTDINNWKYAAETCAKRIINPNLLIVIEG-EAVPKDDVTWTSKSSSYSTWGGNLL 828
QY 218 QGAGQYPPVWLVNPNRLVYSAHDYATSVYPQTWFSPTFPNN--MPGIWNKNGYLFNQ 274
```

QY 125 ALWYTSVSEATWISDQALAQRYKGNPTVYVFDLHNEPHDPACWCGDPSIDWRLAAER 184
 Db 144 ELWYDTSYSEKQWIDDLRFVAHRYANVPGVIGLDVKNPHGRATWGTGDKPTDNTAVEH 203
 QY 195 AGNAVLNVNPNLLIFVEGVQSVNGDS----YMWGNLQAGQYPPVNLNVP-NRLVYSAHD 239
 Db 204 AAAALEAPKWLICGVEGIGENPSCSTIGHFWGENLEPMDCTP--LKVPADHLLLPVH 261
 QY 240 YATSVYPTWFSDDPTFPNNMPCIMNKMWGLFNQNIAPVWLGEFGTTL--QSTTDQWLK 297
 Db 262 YGPDVYVQYFNSPDPFPNNMAAIWDXHGFHAKAGYA-MAIGEGGKYGEGDPRDVAQON 320
 QY 298 TLVQLRPTAQVG-ADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
 Db 321 ALVDYL---ISIGVTDTF---YMSINANSSTGGILFRDDMNHV 357

RESULT 8

Q9V052 PRELIMINARY; PRT; 514 AA.
 AC Q9V052;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Endoglucanase
 GN CELB-LIKE OR PYRAB09400 OR PAS0632.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS / Orsay;
 RA Hellig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ248285; CAB49854.1; --
 DR PIR; E75142; E75142.
 DR HSGP; P54383; IECE.
 DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001547; Glyco hydro_5.
 DR Pfam; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Complete proteome.
 SQ SEQUENCE 514 AA; 59980 MW; CF80CAB1525337C5 CRC64;

Query Match 34.7%; Score 688.5; DB 17; Length 514;
 Best Local Similarity 41.6%; Pred. No. 2.3e-46;
 Matches 148; Conservative 58; Mismatches 121; Indels 29; Gaps 10;
 QY 7 HTSGREILDANNV-----PVRIAGINWFGPCTCNVYVHGLMSRDSYRMLDQIKSLGYNT 60
 Db 30 YTAENGIFVQNTTGEKKPLYLHGVSMFGFELKHVYVGLDKRNVKDKILDKVRLGFNA 89
 QY 61 IRLPYSDDILKPGTMFN--SINFQMNQDLQGLTSQVMDKIVAYAGQIGLRIILDRHP 118
 Db 90 IRLPFCSEIRPDRPSPERIN--ELNPDCLKNLTSLSEIMEKILEYANSIGYILLDVHRI 148
 QY 119 DCSQSALWYTSVSEATWISDQALAQRYKGNPTVYVFDLHNEPHDPACWCGDPSIDW 178
 Db 149 GCEIEPLWYTENYSESYQIKDWIFLAKRFKYPNVIGADIKNEPGEAGWGTDGER-DF 207
 QY 179 RLAAERAGNAVLNVNPNLLIFVEGVQ-----SYNGDSYN--WGNLQAGQYPPVNL 227
 Db 208 RLFAEKVGREILKVAHPLIFVEGTQYTHVNPIDEIIEKGGWTFNGENLGMVKYDIPVRL 267
 QY 228 NVNRLVYSAHDYATSVYPTWFSDDPTFPNNMPCIMNKMWGLFNQNIAPVWLGEFGTTL 287
 Db 268 -PRGVVYSPHVYGPSYVMDYFKSPDPFPNNMPIIETHFGYLTDLNLTIV-ICEWGGNY 325

QY 288 QSTTDQWLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTVDTVK 343
 Db 326 EG-LDKVQWDAFVKWLJKKKIY-----NFFYWCNLPESGDTGGILPDDKNTVWEK 375

RESULT 9

Q97KK6 PRELIMINARY; PRT; 482 AA.
 AC Q97KK6;
 DT 01-OCT-2001 (TREMELrel. 18, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Possible non-processive endoglucanase family 5, secreted, Cella
 DE homolog secreted, dockerin domain.
 GN CAC0912.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Daly M.J.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucalle P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AB007607; AAK78888.1; --
 DR PIR; E97012; E97012.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR002048; EF-hand_5.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR PROSITE; PS00018; EF HAND; 2.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Complete proteome.
 SQ SEQUENCE 482 AA; 54109 MW; 13B2E29197D59A5F CRC64;

Query Match 34.3%; Score 681; DB 16; Length 482;
 Best Local Similarity 37.6%; Pred. No. 8.3e-46;
 Matches 140; Conservative 71; Mismatches 127; Indels 34; Gaps 9;
 QY 5 YWHTSGREILDANNVVPRIAGINWFGPCTCNVYVHGLMSRDSYRMLDQIKSLGYNTIRLP 64
 Db 24 YLHDSGSKLLDDYGGQVQVMTGIAWFGLETPNYCFGLWANRLDNLINIVADNGFWTLRVP 83
 QY 65 YSDIL---KPT--MPNSINPRQNDLQGLTSQVMDKIVAYAGQIGLRIILDRHPD 119
 Db 84 LSELVNLWQVQGYPTPSIN-DYISPELKGQNSQLDDVIAYSKVGKVMVMDHRIE 142
 QY 120 CSQSALWYTSVSEATWISDQALAQRYKGNPTVYVFDLHNEPHDPAC-----WCGG 172
 Db 143 SGQATATWYTSKYTTDDYEKCQYLADRYKDDTVIAADI FNEPHGKAYRAETSAKNDT 202
 QY 173 DPSIDWRLAAERAGNAVLNVNPNLLIFVEGVQSY-----NGDSY---WNGNLOQ 219
 Db 203 TDEDNRVYEAERKVKKILDPINPKMLIVVEGVETPKGTAAGSTNPDYVGGWGNLGR 262
 QY 220 AQYPPVNL-NVNPRLVYSAHDYATSVYPTWFSDDPTFPNNMPCIMNKMWGLFNQNIAP 277
 Db 263 VKDYFPVLDAPYKNQVYVSPHYDYGCVSQDTWFDGFTQSLINDIWRPSWFIQEKNIAP 322
 QY 278 VWLGEFGTTLQSTTDQWLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQ 337
 Db 323 LLIGEWGGNMGDKGKQEWQNTDMAXLI-----SDNKGHTFWCLNANSGDTGGILEYDFK 376


```
Db 262 YGPDVFQVQSFNDNFNNPMPALWDRHFGQFAGSH--ALLLGEFGGKYCEGDARDKWQD 319
QY 298 TLVQLRPTAQGADSFQWTFWNPDSGDTGGILKDDWQTV 339
Db 320 ALVKYLR---SGIN--EGFYWNPNSGDTGGILRDWTSV 356

RESULT 6
OS8925 PRELIMINARY; PRT; 458 AA.
ID OS8925
AC OS8925;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 458AA long hypothetical endo-1,4-beta-glucanase.
GN Phil71.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RC STRAIN=OT3.
RX MEDLINE=983744137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hatakeyama Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama M., Ohfuku Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000005; BAA30271.1; -.
DR PIR; E71059; E71059.
DR HSP; P54583; IECE.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR InterPro; IPR006162; Plantne_S.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
KW Complete proteome.
SQ SEQUENCE 458 AA; 51930 MW; CAE48AD3A8EE654 CRC64;

Query Match 38.4%; Score 762; DB 17; Length 458;
Best Local Similarity 45.0%; Pred. No. 2.8e-52;
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSLDQIKSLGYNTRILPYSD 67
Db 54 TSGEE-----TPHLFGVKNWFGFETPNVHVGKLNKRWEDMLQIKSLGFNAIRLPFCT 107
QY 68 DILKPGTNPNSINFRMNQDLQGLTSIQVMDKIVAYAGQIGLRIILDRHRPDCSGSALW 127
Db 108 ESVKPGTQPIDYSK-NPDLRGLDSIQIMEKIKKAGDLGIFVLLDHYRIGCTHIEPLW 166
QY 128 YTSVSSEATWSDLOALQAKQKGNFTVVGEDLHNEPHD-----PACWGCGDPS 175
Db 167 YTEFSEEDFTNTWIEVAKRGKTNVIGADLKNKPHSVTSPPAAYTDGTGATWGNFPA 226
QY 176 IDWRLAERAGNVLNPNLLIFVEGVQSYN-----GDSYWGMLQAGQYVYV 226
Db 227 TDWNLAEIRICKAILKVAHPHLIFVEGTQFTNPXTDSSYKGYNAWGNLMAVKDYPV- 285
QY 227 LNVP-NRLVYSAHYATSVPQTWFSDEPT-PPNMPGINKWVLFNQNIAPVWLGEGF 284
Db 286 -NLPRNKLVSPHYVGPVYNQYFGPAKGPDPNLPDIWHFFGVKLELGYSVVIGEGF 344
QY 285 TTL---QSTTQDTWLTQLVQLRPTAQYGADSFQWTFWNPDSGDTGGILKDDWQTV 339
Db 345 KYGHGGPRDVIWQNKLVDMW--IENKFCDF--YSWNPDSGDTGGILQDDWTI 397
```

```
RESULT 7
Q9PF60 PRELIMINARY; PRT; 592 AA.
ID Q9PF60
AC Q9PF60;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN XP0818.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Farga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais W.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montei-Vitorcello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.R., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshuko M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.B.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
DR EMBL; AE003921; BAF83628.1; -.
DR FIR; B82759; B82759.
DR HSP; P54583; IECE.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac cellose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD_II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 592 AA; 59967 MW; 9846DA4EA3B5C89E CRC64;
```

```
Query Match 35.0%; Score 695.5; DB 16; Length 592;
Best Local Similarity 42.6%; Pred. No. 7.7e-47;
Matches 146; Conservative 52; Mismatches 126; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSLDQIKSLGYNTRILP 64
Db 26 YSISHGKVIDDKGN-QIQLKGSWFGFETTNVHVGKLNKRWKFEITQSGMLNAVRP 84
QY 65 YSDILKPGTNPNSINFRMNQDLQGLTSIQVMDKIVAYAGQIGLRIILDRHRPDCSGS 124
Db 85 FCPASLNSNTSPSSIDYNN-NPDLQGLSLQIMDKVVKLSDRGIYVLMDDHTPDCAIS 143
```

```

Db 165 SELWYTSYPSRWSIDWRMLAKRYASDPTVIGVDLHNEPHCAATWGTGAATTDWRAAAE 224
QY 184 RAGNAVLGVNPNLLIFVEGV-QSYNGDSYWMGNLQAGQPVVVLNPNRLVVSADHYAT 242
Db 225 RGNNAVLAENPKLLVVEGIDHQADGTCTWGGALDSATASVELTVANRVVYSPHDYPS 284
QY 243 SVYPTWFSDDPTFPNNPFGIWNKNGYLFNQNIAPVMLGEGFTLQSTTDDTWTUKTLVQY 302
Db 285 TIYGQWFSASNYPTNLPGLIADHWGYLAKDIAPVLVGEFTGKLETASDKQWLNLTLYG 344
QY 303 LRPTAQYCADSQFTWFSWSPDSGDTGGILKDDQCTVDVTKDGYLAPI 350
Db 345 LSSIT---GISS---SFWAFNPDSDGTGIVASDWVTPQAKLDALAPI 386

RESULT 4
Q9EVR6 PRELIMINARY; PRT; 518 AA.
AC Q9EVR6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Extracellular endoglucanase (ENGXCA protein) precursor.
GN ENGXCA.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRLL-B1459;
RA Schroeter K., Pushler A., Becker A.;
RT "engXCA major extracellular endoglucanase.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304415; CAC18529.1; --
DR HSSP; P54583; 1EC6.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR Pfam; PF00150; Cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT CHAIN 1 25 POTENTIAL.
FT CHAIN 26 518 EXTRACELLULAR ENDOGLUCANASE (ENGXCA
PROTEIN)
SQ SEQUENCE 518 AA; 55588 MW; 9A0FF678B3B712C CRC64;

Query Match 40.5%; Score 803.5; DB 2; Length 518;
Best Local Similarity 46.9%; Pred. No. 1.6e-55;
Matches 161; Conservative 61; Mismatches 102; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRYSMLDQIKSLGYNTRILP 64
Db 26 YSINNSRQIVDDSGKVLQKGVNVFGFTGNVHGLWLRNWKDMIVQMGLGFNAVRLP 85
QY 65 YSDILKPGTNPNSINFRQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHPCDGSQS 124
Db 86 FCPATLRSDTWPASIDYSR-NADLQGLTSLQILDKVIAEFNARGWVYLLDHTPDGAGIS 144
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLNHEPHDPCWCGDPSIDWRLAER 184
Db 145 ELWYTGYSYEAQWLDDLRFRVANNRYKNVPSVIGVLDLKNPEHGAATWGTGNAATDWNKAAE 204
QY 185 AGNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQPVVVLNPN-NRLVVSADH 238
Db 205 GSAVLAVAPKWLIAVGGITDNPVCSNG-GIFWGNLQPLACTP--LNIPANRLLLAPH 261
QY 239 DYATSVPTQWFSDDPTFPNNPFGIWNKNGYLFNQNIAPVMLGEGFTL--QSTTDTWL 296
Db 262 VYGPDVVFQSYFNSFPNNPFAIWERHFGQFAGTH--ALLGFGGKGEGDARDKTWQ 319
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDQV 339

```

```

Db 320 DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

RESULT 5
Q8PPS3 PRELIMINARY; PRT; 474 AA.
AC Q8PPS3;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Cellulase.
GN ENGXCA OR XAC0612.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida M.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camnava F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorzy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura E.C., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB011689; AM35501.1; --
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00350; Cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 51294 MW; 379B84D63CASA31B CRC64;

Query Match 40.4%; Score 802.5; DB 16; Length 474;
Best Local Similarity 46.2%; Pred. No. 1.7e-55;
Matches 159; Conservative 63; Mismatches 102; Indels 19; Gaps 9;

QY 6 WHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRYSMLDQIKSLGYNTRILP 65
Db 26 YSINNSRQIVDDSGKVLQKGVNVFGFTGNVHGLWLRNWKDMIVQMGLGFNAVRLP 85
QY 66 YSDILKPGTNPNSINFRQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHPCDGSOSA 125
Db 86 FCPATLRSDTWPASIDYSR-NADLQGLTSLQILDKVIAEFNARGWVYLLDHTPDGAGIS 144
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLNHEPHDPCWCGDPSIDWRLAER 185
Db 145 ELWYTGYSYEAQWLDDLRFRVANNRYKNVPSVIGVLDLKNPEHGAATWGTGNAATDWNKAAE 204
QY 186 AGNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQPVVVLNPN-NRLVVSADH 239
Db 205 GSAVLAVAPKWLIAVGGITDNPVCSNG-GIFWGNLQPLACTP--LNIPANRLLLAPH 261
QY 240 YATSVPTQWFSDDPTFPNNPFGIWNKNGYLFNQNIAPVMLGEGFTL--QSTTDTWLK 297

```

```
QY 124 SALWYTSVSEATWISDLQALAQRYKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 156 SELWYTSQYPSRWISDWMLAERYKXNPTVIGADLHNEPHDQAGWGTVDWRLAAQ 215
QY 184 RAGNAVLSVNPMLLIFVEGVQ---SYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDY 240
Db 216 RAGNAVLSVNPMLLIFVEGVQYDHNKNSQYWMGNLQAGQYPPVVLNPNRLVYSAHDY 275
QY 241 ATSVYPTQWFSDFTPNNPGLNKNWGYLFNQNTAPVWLGBFG--TTLQSTTDOTWLT 298
Db 276 GPCVSSQPFNDSTFPSNLPAIWDQWGYISKQNIAPVLVGFGRNVDLSPEGKQWNA 335
QY 299 LVQYLPTAQYAGDSFQWTFWSNPDGDTGGLKDDWQTVTDVKGVLAPIKSIFDPV 358
Db 336 LVYDI-----GANNLYFIWLSNPNSTGDTGGLLDDWTTWRPKQDML-----SRIMKPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
AC Q9K5C7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OG Plasmid pCMI.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Meletus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RV [2]
RW SEQUENCE OF 1-614 FROM N.A.
RX STRAIN=NCPPB 382;
RY Meletus D.;
RZ Meletus D.;
RA Jah H.; Drier J.; Meletus D.; Bahr R.; Eichenlaub R.J.;
RT "the endo-beta-1,4-glucanase Cella of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato.";
RL Mol. Plant Microbe Interact. 13:703-714 (2000).
RV [3]
RW SEQUENCE FROM N.A.
RX STRAIN=NCPPB 382;
RY Jah H.;
RZ Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RA Jah H.;
RL EMBL; X62582; CAA44467.2; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellulase.
DR InterPro; IPR001547; Expan_Lol_pi_C.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Plasmid; Signal.
FT SIGNAL 1 66 POTENTIAL.
FT CHAIN 67 >745 CELLULASE.
FT NON TER 745 745
SQ SEQUENCE 745 AA; 77886 MW; 5945015EB8D4F27C CRC64;

Query Match 45.8%; Score 909.5; DB 2; Length 745;
Best Local Similarity 50.3%; Pred. No. 1e-63;
Matches 175; Conservative 51; Mismatches 111; Indels 11; Gaps 4;
```

```
QY 4 GYWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRL 63
Db 70 GMLHTAGKLVITAGCAPYTRGIANFQWETSSCAPHGLDITILAGGQWQH-KONGFTTVRL 129
QY 64 PYSDDLKPGTWPNSINFRQMQDLOGLTSLQWMDKIVAYAGQIGRLIILDRHRPDCSQ 123
Db 130 PFSNOCL-----AASGVTGVSDPSLAGLTPLQVMDHVVASAKAGLDVILDOHRPDSGGQ 185
QY 124 SALWYTSVSEATWISDLQALAQRYKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 186 SELWYTSQYPSRWISDWMLAERYKXNPTVIGADLHNEPHDQAGWGTVDWRLAAE 245
QY 184 RAGNAVLSVNPMLLIFVEGV---QSYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDYAT 242
Db 246 RGNNAVLAENPNLLVVEGIDHEADGSGTWMGALGLVGNAPVRLSVANRVVYSPHDYPS 305
QY 243 SVYPTQWFSDFTPNNPGLNKNWGYLFNQNTAPVWLGBFGTTLQSTTDOTWLTWLTQVY 302
Db 306 TIYQSWFASNYFANLPGIWDHAGWGLAKDIAVLVGFEGTKFKETISDKWNLTVGY 365
QY 303 LRPTAQYAGDSFQWTFWSNPDGDTGGLKDDWQTVTDVKGGYLAPI 350
Db 366 LSST---GISS---SFWAFNPNSGDTGIVKSDWVTPQAKLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
AC Q9AF65;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Cellulase Cella.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Laine M.; Haapalainen M.; Wahlroos T.; Kankare K.; Nissinen R.;
RA Kassuwi S.; Metzler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain.";
RL Physiol. Mol. Plant Pathol. 0:0-0 (2001).
DR EMBL; AY007311; AAKJ6222.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellulase.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 727 AA; 75811 MW; 96DF9B664873985D CRC64;

Query Match 45.4%; Score 902.5; DB 2; Length 727;
Best Local Similarity 50.3%; Pred. No. 3.5e-63;
Matches 175; Conservative 48; Mismatches 114; Indels 11; Gaps 4;
```

```
QY 4 GYWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRL 63
Db 49 GMLHTAGKLVITAGCAPYTRGIANFQWETSSCAPHGLDITILAGGQWQH-KONGFTTVRL 108
QY 64 PYSDDLKPGTWPNSINFRQMQDLOGLTSLQWMDKIVAYAGQIGRLIILDRHRPDCSQ 123
Db 109 PFSNOCL-----AASGVTGVSDPSLAGLTPLQVMDHVVASAKAGLDVILDOHRPDSGGQ 164
QY 124 SALWYTSVSEATWISDLQALAQRYKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 ; Search time 36.3333 Seconds
(without alignment)
3108.867 Million cell updates/sec

Title: US-09-997-504A-14
Perfect score: 1986
Sequence: 1 AGGQVWHTSGREILDANNVP.....VDTVKDGLYAPIKSSIFDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1049	52.8	397	2 Q8RP23	Q8RP23 paenibacill
2	909.5	45.8	745	2 Q9K5C7	Q9K5C7 clavibacter
3	902.5	45.4	727	2 Q9AF65	Q9AF65 clavibacter
4	803.5	40.5	518	2 Q9EVR6	Q9EVR6 xanthomonas
5	802.5	40.4	474	16 Q8PFS3	Q8PFS3 xanthomonas
6	762	38.4	458	17 Q58925	Q58925 pyrococcus
7	695.5	35.0	592	15 Q3P260	Q3P260 xylella fas
8	685.5	34.7	514	17 Q5V052	Q5V052 pyrococcus
9	681	34.3	482	16 Q97KK6	Q97KK6 clostridium
10	656.5	33.1	614	16 Q87AH4	Q87AH4 xylella fas
11	629.5	31.7	1000	2 Q24820	Q24820 thermophili
12	623	31.4	1426	2 Q9X3P6	Q9X3P6 caldicellul
13	621.5	31.3	660	2 Q9L3J2	Q9L3J2 clostridium
14	602.5	30.3	534	2 Q9EY00	Q9EY00 clostridium
15	583	29.4	616	2 Q9X2N2	Q9X2N2 thermomonos
16	524.5	26.4	425	2 Q93Q07	Q93Q07 thermus cal

17	492	24.8	574	2 Q9S3V3	Q9S3V3 cellulomona
18	490	24.7	341	2 P96310	P96310 anaerocellu
19	393	19.8	630	2 Q8RJY7	Q8RJY7 stigmatella
20	324	16.3	565	16 Q9PF69	Q9PF69 xylella fas
21	311.5	15.7	569	16 Q87AG9	Q87AG9 xylella fas
22	272	13.7	508	10 Q9LTM8	Q9LTM8 arabidopsis
23	271.5	13.7	590	16 Q8P513	Q8P513 xanthomonas
24	257.5	13.0	522	10 Q9SAB6	Q9SAB6 arabidopsis
25	210	10.6	526	10 Q9LFS2	Q9LFS2 arabidopsis
26	206.5	10.4	488	10 Q9LFE7	Q9LFE7 arabidopsis
27	194.5	9.8	354	16 Q8PEF2	Q8PEF2 xanthomonas
28	190.5	9.6	555	10 Q7XUQ4	Q7XUQ4 cryza sativ
29	188.5	9.5	561	10 Q9LTM0	Q9LTM0 arabidopsis
30	184.5	9.3	437	3 Q99036	Q99036 trichoderma
31	176.5	8.9	377	3 Q00012	Q00012 aspergillus
32	170.5	8.6	694	2 Q9Z187	Q9Z187 bacillus st
33	169	8.5	439	3 Q92401	Q92401 agaricus bi
34	168	8.5	439	3 Q9F893	Q9F893 agaricus bi
35	166.5	8.4	357	16 Q8PRD3	Q8PRD3 xanthomonas
36	166	8.4	377	16 Q8PRD5	Q8PRD5 xanthomonas
37	164.5	8.3	558	10 Q7XUQ6	Q7XUQ6 cryza sativ
38	163	8.2	364	2 Q9F0G8	Q9F0G8 rhizobium m
39	160.5	8.1	516	2 Q66185	Q66185 bacillus ci
40	160	8.1	550	10 Q7XUQ5	Q7XUQ5 cryza sativ
41	154.5	7.8	722	16 Q97L56	Q97L56 clostridium
42	153.5	7.7	363	2 Q07652	Q07652 cellvibro
43	153.5	7.7	1449	10 Q8RU51	Q8RU51 cryza sativ
44	153	7.7	357	2 Q59441	Q59441 fibrobacter
45	151.5	7.6	356	16 Q9PA12	Q9PA12 xylella fas

ALIGNMENTS

RESULT 1

Q8RP23 PRELIMINARY; PRT; 397 AA.

AC Q8RP23; 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Endo-beta-1,4-glucanase

OS Paenibacillus sp. KCTC8948P.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI_TaxID=109199;

RN [1]

RC SEQUENCE FROM N.A.

RC STRAIN=KTC 8848P;

RA Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;

RT "Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-

RT expression with Endomyces fibuliger beta-glucosidase gene in

RT Saccharomyces cerevisiae."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF345984; AAL8749.1; ..

DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00150; cellulase; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

SQ SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;

Query Match 52.8%; Score 1049; DB 2; Length 397;
Best Local Similarity 54.7%; Pred. No. 3.3e-75;
Matches 197; Conservative 44; Mismatches 103; Indels 16; Gaps 5;

QY	4	GWHTSGREILDANNVPVRIAGINWPGTCTVYVHGLWSRDMQIKSLGNTRL 63
DB	37	GYHTQGNKIVDETGEAAFNGLWFTPNYTLHGLWSRMDMLDQVKEGYNLRL 96
QY	64	PYSDDTLKPGTWPNINFRQNDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRDCSGQ 123
DB	97	PSYNQLFDSSSRADSIDYK-NPDLVGLTPTQIMDKLEKAGQGIQLDLDRHRCGSGQ 155

This Page Blank (uspto)

Db 402 QVWDSLSRGFTVITINSHDD-----WIKEDYNGNIERFEKIW-----EQIAERFKNKS 450
 QY 153 TVGFDFLHNEPHDPACGCGPDSIDWLAARAGNAVLNVNPNLLIFVEGVQSYNGDSYV 212
 Db 451 ENLIFEIMNEFFG-----NITDEQID--DMNSRIILKIIRKINPTIVII-----GGYV 497
 QY 213 WGNLQAGAGVPPVVLN--VPN--RLVSAHDYATSVYQVWFSDDP--TFPNNMPCIVN--- 264
 Db 498 -----NSYNTLVNIKIPDDPYLIGTFHYV-----DPYBFTHKWRGTWGTQOE 538
 QY 265 -----KNGVLYFNQNIAPVWLGEGTTLQSTTDQTLKTLVQLRPTAQYAD 312
 Db 539 DMDTVRVDFVKWS-----DRNIPVTFGEF--AMVAYADTSRVKYDFISDAAL--ER 591
 QY 313 SFQWTFMSWNPDSGDTGILKD-----DMQVDTVKDGYLAPIKSSIFDP 357
 Db 592 GFACSVW-----DNGVFGSLDNDMAIYNRDTRTFDT-----EILNALFNP 631

RESULT 15
 GUNB_RUMAL
 ID GUNB_RUMAL STANDARD; PRT; 409 AA.
 AC P23661.
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase) (EGB).
 GN CELB.
 OS Ruminococcus albus.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
 OC Ruminococcus.
 OX NCBI_TaxID=1264;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SY3;
 RX MEDLINE=91066833; PubMed=2250649;
 RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;
 RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
 RT genes celA and celB."
 RL Mol. Gen. Genet. 223:217-223(1990).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54932; CAA38693.1; -.
 CC PIR; S12018; S12018.
 CC HSSP; P17901; 1EDG.
 CC InterPro; IPR001547; Glyco hydro. 5.
 CC InterPro; IPR000437; Prok lipoprot_s.
 CC Pfam; PF00150; cellulase; 1.
 CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
 CC DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
 CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 CC SIGNAL 1 21 POTENTIAL.
 CC FT CHAIN 22 409 ENDOGLUCANASE B.
 CC FT ACT_SITE 212 PROTON DONOR (BY SIMILARITY).
 CC FT ACT_SITE 332 332 NUCLEOPHILE (BY SIMILARITY).
 CC SQ SEQUENCE 409 AA; 45523 MW; 119081DFA3BFDD54 CRC64;

Query Match 6.7%; Score 134; DB 1; Length 409;
 Best Local Similarity 24.1%; Pred. No. 0.0027;
 Matches 78; Conservative 44; Mismatches 119; Indels 82; Gaps 16;

QY 49 MLDQIKSLGYNTIRLPYS-----DDLKPGTNPNSINFRQMQDLOGLTSLQVMDKI 100
 Db 112 MIDMLPEAGFNVLRIPIVSWGNHLIDNNYTIIDPAWMDR-----VQEI 152
 QY 101 VAYAGIGLRILDRHRPDCSQSALWYTSVSEATW-ISDLQA-----LAQRYKGNPTV 155
 Db 153 VNYGIDDDGMVILNTHHEE-----WYMPKPEKDGIDIELKAIWSQIADRFKGYDEHL 205
 QY 156 GFDLHNEPHDPACWCGDPDSIDWRLAERA-----GNAVLSVNPMLLIFVEGVQS---YN 207
 Db 206 IFEGLNEP-----RLRGEAEWTGTSEAREIINEYEKAFVETVRASGGNN 250
 QY 208 GDSYWW-----CGNLQAGAGQVPLNVNPNRLVYSADYAT-SVYPTWFSQPTFPNN 258
 Db 251 GDRCLMITGYAASSGYNLSAIELP---EDSKLLISVHAYLPYSFALDTGTDKYDPE 307
 QY 259 --MPGIWKNKNGYLEFNQNIAPVWLGEFGTTLQSTTDDQTLKTLVQLRPTAQYGADSPQW 316
 Db 308 TAIPTLP-ESSLNELFISRDIPVIVGEFGSMKNKONIDDR-VKCLDDYLGNAAKYDIPCVWM 365
 QY 317 TFWSWNPDSGDTGGIL---KDDW 336
 Db 366 DNYA-RIGNGENFGLLRQYDW 387

Search completed: August 2, 2004, 16:34:57
 Job time : 9.33333 secs

```
CC      WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC      MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC      -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC      hydrolases).
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; D90341; BAA14354.1; -.
CC      HSSP; P17901; 1EDG.
CC      InterPro; IPR005087; CBM_11.
CC      InterPro; IPR002105; Dockerin_1.
CC      InterPro; IPR002048; EF-hand_1.
CC      InterPro; IPR001547; Glyco_hydro_5.
CC      Pfam; PF03425; CBM_11; 1.
CC      Pfam; PF00150; cellulase; 1.
CC      Pfam; PF00404; Dockerin_1; 2.
CC      PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC      PROSITE; PS00448; CLOS CELLULOSE RPT; 1.
CC      PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
CC      Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC      SIGNAL 1 24
CC      CHAIN 25 584
CC      DOMAIN 25 328
CC      DOMAIN 329 353
CC      DOMAIN 354 584
CC      DOMAIN 530 552
CC      REPEAT 530 552
CC      REPEAT 562 584
CC      ACT_SITE 159 159
CC      ACT_SITE 264 264
CC      ACT_SITE 264 264
CC      SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;
CC
CC      Query Match 7.2%; Score 143.5; DB 1; Length 584;
CC      Best Local Similarity 21.1%; Pred. No. 0.00074;
CC      Matches 71; Conservative 53; Mismatches 140; Indels 63; Gaps 17;
CC
CC      QY 21 VRIAGINWGFETCNVYVHGLSRDYS-MLDQIKSLGNTIRLPYSDDLKPGTWPNIS 79
CC      Db 32 VKKMGIGMLNLTGTFDPTGSGSKAAQYVYFDDFKQAGFHVIRPQDHTLANSFYTV 91
CC      QY 80 NFRQMNDLQGLSLQVMDKIVAVAGIGILRIILDRHRPCSGQSALW---YTSSVS-- 133
CC      Db 92 DSNFLNR-----IETVDSLSRGFVTVINSHHD-----TWLMDNYSQNIQRP 134
CC      QY 134 EATWISDLQALQRYKGNPTVVGFDLHNEPDPACWCGDPSIDWLAERAGNAVLSVN 193
CC      Db 135 EKIW---EQIAQRFKGSNLFVETILNEPHG---NITDSQIN--DMNKRILNIIRKTN 184
CC      QY 194 PNLLIFVEGVQSGVNGDSYWGNNLQAGQYVVLNPNLVNRYSAHDYATSVYPTQWFSDP 253
CC      Db 185 PTRNVII-----GAGYWNYSNLSOLEIP---NDPN-LIATFYHDPVSFTHQMGWTW 233
CC      QY 254 TFPNNMPEGI---WN--KNWGYLFNQNIAPVWLGEFGTTLQSTTDQTLVQLRPTA- 307
CC      Db 234 GTKNDMDALIAVFNHVKWS---DKNNIPVYLGEYG--VMGHSRDSAVKWFDFVSDQAI 288
CC      QY 308 QYGADSFQWTFWSPNPSDGTGILKD 334
CC      Db 289 SHGFSCGAW-----DNGYFGVDND 308
CC
CC      RESULT 14
CC      GUNH_CLOTH STANDARD; PRT; 900 AA.
CC      AC P16218;
CC      DT 01-APR-1990 (Rel. 14, Created)
CC      DT 01-APR-1990 (Rel. 14, Last sequence update)
```

```
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase)
DE (Cellulase H).
GN CELH.
OS Clostridium thermoCELLUM.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_taxid=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=90323606; PubMed=2197182;
RA Yaguee E., Beguin P., Aubert J.-P.;
RT "Nucleotide sequence and deletion analysis of the cellulase-encoding
RL gene celiH of Clostridium thermoCELLUM.";
CC -!- GENE 89:61-67(1990).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 26 OF
CC GLYCOSYL HYDROLASES.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; M31903; AAA23225.1; -.
CC      PIR; JH0157; JH0157.
CC      HSSP; P07985; 1CEC.
CC      InterPro; IPR005087; CBM_11.
CC      InterPro; IPR002105; Dockerin_1.
CC      InterPro; IPR002048; EF-hand_1.
CC      InterPro; IPR002048; EF-hand_1.
CC      Pfam; PF03425; CBM_11; 1.
CC      Pfam; PF00150; cellulase; 1.
CC      Pfam; PF00404; Dockerin_1; 2.
CC      PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC      PROSITE; PS00448; CLOS CELLULOSE RPT; 2.
CC      PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
CC      Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC      SIGNAL 1 44
CC      CHAIN 45 900
CC      DOMAIN 45 630
CC      DOMAIN 631 654
CC      DOMAIN 655 900
CC      ACT_SITE 460 460
CC      ACT_SITE 565 565
CC      DOMAIN 833 895
CC      REPEAT 833 856
CC      REPEAT 872 895
CC      SEQUENCE 900 AA; 102415 MW; 973AFB1954FC246B CRC64;
CC
CC      Query Match 6.9%; Score 137; DB 1; Length 900;
CC      Best Local Similarity 21.1%; Pred. No. 0.0041;
CC      Matches 74; Conservative 51; Mismatches 122; Indels 104; Gaps 21;
CC
CC      QY 40 GLWSD-YRSMLDQIKSLGNTIRLPYSDDLKPGTWPNISINFRQMNDLQGLSLQVMD 98
CC      Db 352 GWSKSKAMEYFDDFKQAGYKKNVIRPVRWDNHTVTRTYPTID-----KAFLDVE 401
CC      QY 99 KIVAVAGQIGILRIILDRHRPCSGQSALW-----YTSSVS--EATWISDLQALQRYKGNP 152
CC      Db 99 KIVAVAGQIGILRIILDRHRPCSGQSALW-----YTSSVS--EATWISDLQALQRYKGNP 152
```

Query Match 7.9%; Score 156.5; DB 1; Length 814;
Best Local Similarity 20.9%; Pred. No. 0.0001;
Matches 81; Conservative 60; Mismatches 133; Indels 113; Gaps 19;

QY 6 WHTSGRILDANNVPVRIAG-----INWFGFETCNVYVHGLWS--RDYRSM 49
DB 32 FQVSGTKLLDASGNELVMRGMRDISAIDLKVKIKGNLGNLTLDAPTETAGNPRTTKAM 91
QY 50 LDQIKSLGYNTIRLPYSDDILKPGTMTN-SINFRQMNQDLQGLTSLQVMDKIYVAGQIG 108
DB 92 IEKREMGFNARVPVFTWD-THIGAPDYKIDELANR-----VEEVNYYVLDCG 140
QY 109 LRIILDRHRPCSCQSALWTSSSEATWISDLQA-----LAQYKGNPT 153
DB 141 MYALINLHH-----DNTWLIPTVANEQRSEKELVKWVEQIATRFKDYDD 184
QY 154 VVGDLNHEPHDPACWCGDPSIDWRLAABRAGNAVLSVNPNNLI FVEGVQSYNGDSYWW 213
DB 185 HLLFETWNEPRE-----VGSP-MEWMGGTYENRDNFRN-----LAVNTIRAS- 228
QY 214 GGN-----LOGAGQYPVVLN---VPN---RLVYSADHYATSVYPQT-----WFS 252
DB 229 GGNNDKRFILVPTAATGLDALNDLVI PNNDSRVIVSIHAYSYPFFAMDVNGTSYWGSD 288
QY 253 ---PTFPNNMGIWNKNGWYLFNQNIAVPWLGERGTTLQSTTDQTLWTKLV-----QYL 305
DB 289 YDKASLTSELDIYNR-----FVKNGRAVVIIGFEG-----TIDKNLSRVAHAHYARE 338
QY 306 TAOQAGDSFQWTFWNPDSGDTGAIL 332
DB 339 AVSRGIATVWDNGYVPGDAETVALL 365

RESULT 12
GUND_RUMAL STANDARD; PRT; 406 AA.
AC P16216;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EG-I).
GN EG I.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.
RC STRAIN=F-40;
RX MEDLINE=90078126; PubMed=2687251;
RA Ohmura K., Kajino T., Kato A., Shimizu S.;
RT "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene."
RL J. Bacteriol. 171:6771-6775 (1989).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M30528; AAA26469.1; -
DR PIR; A43722; A43722.
DR HSSP; P17901; 1EDG.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 43
FT CHAIN 44 406 ENDOGLUCANASE I.
FT ACT SITE 210 210 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 330 330 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45390 MW; 2E0172437B14FEA8 CRC64;
Query Match 7.4%; Score 146.5; DB 1; Length 406;
Best Local Similarity 22.9%; Pred. No. 0.00027;
Matches 88; Conservative 49; Mismatches 152; Indels 95; Gaps 18;

QY 15 DANNYVRIAGIN-----WFGFETCNVYVHGLWSR-----DYR 47
DB 49 ETENFVFSQTHNTDNTMTVTSKDLVAKWTNGNLTMDATAQGLGSEVSLPLKVTNK 108
QY 48 SMLDQIKSLGYNTIRLPYSDDILKPGTMTN-SINFRQMNQDLQGLTSLQVMDKIYVAGQI 107
DB 109 YIMDLPEAGFNVLPIVS-----WGNHILDDKYTSDPAWMDRVQ---EIVNYGIDN 157
QY 108 GURIILDRHRPCSCQSALWTSSVSE-----ATWISDLQALAQRYKGNPTVVGFD 158
DB 158 GLYVILNTHHEB-----WYMPKPEKDGIEEIKAVMAQ---IADRFGYDEHLIFE 206
QY 159 LHNEPHDPACWCGDPSIDWRLAABRAGNAVLSVNPNNLI FVEGVQSYNGDS 210
DB 207 GUNEP-----RLRGEAEWGTSEAREIINEYKAFVETVRASGNGNDR 251
QY 211 YWNGNLOGAGQYP-VVLNVP---NRLVYSADHYAT-SVYPQTWFSPTFPNN--MFGI 262
DB 252 CLMTGYAASSAYNNLSAIELPEPDSKLIISVHAYLPYSFALDTKGTDKYDPEITAIPEL 311
QY 263 WKNWGYLFNQNIAVPWLGERGTTLQSTTDQTLWTKLVQYLRPTAQYAGDSFQWTFWSN 322
DB 312 P-EHLNELFISKGIPIVGEFGPTMKNKENTEDR-VKLEDYLAARAKYDIPCWWDNYA-R 368
QY 323 PDSGDTGGILKD---DWQTVDTVK 343
DB 369 ICGNENFGLMNRADLEWFFDLIE 392

RESULT 13
GUND_CLOCE STANDARD; PRT; 584 AA.
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)
DE (Cellulase D) (EGCCD).
GN CELCCD.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RX MEDLINE=92009193; PubMed=1916275;
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RL Gene 104:33-38 (1991).
CC -!- FUNCTION: The biological conversion of cellulose to glucose
generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellulohydrolases that cut the disaccharide cellobiose
from the nonreducing end of the cellulose polymer chain; (3)
CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other
short cellobio-oligosaccharides to glucose.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PATHWAY: Cellulose degradation.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS


```

[1]
RN RC SEQUENCE FROM N.A.
RP RC STRAIN=AW;
RX RC MEDLINE=92138626; PubMed=1735723;
RA RA Huang J., Schell M.A.;
RT "Role of the two-component leader sequence and mature amino acid
RT sequences in extracellular export of endoglucanase EGL from
RT Pseudomonas solanacearum.";
RL J. Bacteriol. 174:1314-1323 (1992).
RN [2]
RP RC SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.
RX RC MEDLINE=89291722; PubMed=2738021;
RA RA Huang J., Sukordhman M., Schell M.A.;
RT "Excretion of the egl gene product of Pseudomonas solanacearum.";
RL J. Bacteriol. 171:3767-3774 (1989).
RN [3]
RP RC PROCESSING.
RX RC MEDLINE=90307678; PubMed=2195024;
RA RA Huang J., Schell M.A.;
RT "Evidence that extracellular export of the endoglucanase encoded by
RT egl of Pseudomonas solanacearum occurs by a two-step process
RT involving a lipoprotein intermediate.";
RL J. Biol. Chem. 265:11628-11632 (1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M84922; AAA61980.1; -.
DR PIR; A42649; A42649.
DR InterPro; IPR001547; Glyco_hydro.5.
DR InterPro; IPR000437; Prok_LipProt_S.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
KW Lipoprotein; Signal; Palmitate.
FT SIGNAL 1 19
FT PROPEP 20 45
FT CHAIN 46 426
FT LIPID 20 20
FT LIPID 20 20
FT ACT_SITE 249 249
FT ACT_SITE 361 361
FT SEQUENCE 426 AA; 45578 MW; 51E13AD4442CF48 CRC64;
SQ
Query Match 8.4%; Score 166; DB 1; Length 426;
Best Local Similarity 23.2%; Pred. No.8.3e-06;
Matches 82; Conservative 54; Mismatches 139; Indels 78; Gaps 18;
QY 25 GINWFGFETCNVYVHGLWSRDY----RSMLDQIKSLGNTIRLPYSDDILKPGTMPNSIN 80
Db 121 GVSILAGAEFGSGSPGYGSIYIPSDSVTYYNKGNVLRLFRWERLQP----- 172
QY 81 FRWNQ--DLQGLTSLQWDXIVAYAGQIGLRITILDRHRPCDQSGSALWY-----TSSVS 133
Db 173 --TLNQVDFANELSRLTGFNVAVTATGQ---TVLLDPH-----NYARYYGNVITGSSAVP 221
QY 134 EATWISDLQAORYKGNPTVVGDLHNEHPDPCWCGDPSIDWRLAABEAGNVLNVN 193
Db 222 NSAYADFWRRLATQFKSNPRVI-LGLMNEPNSM-----PTEQLWSCANAEIAIRSAN 273
QY 194 PNLLIFVEGVQSYNGDSYVWNGNTLQGAGQYEW--LNVP-NRLVYSADHY----- 240

```


DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (EGB) (Endo-1,4-beta-glucanase)
 DE (Cellulase B).
 GN CELB.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10862;
 RX MEDLINE=86148508; PubMed=34531202;
 RA "Grepinet O., Beguin P.;
 RT "Sequence of the cellulase gene of Clostridium thermocellum coding
 RT for endoglucanase B";
 RL Nucleic Acids Res. 14:1791-1799 (1986).
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03592; CAA27266.1; -;
 DR PIR; A23512; CZCLEM.
 DR HSP; P54583; LECE.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 27 OR 31.
 FT CHAIN 28 563 ENDOGLUCANASE B.
 FT ACT_SITE 204 204 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 502 557 2 X 24 AA APPROXIMATE REPEATS.
 FT REPEAT 502 526 1.
 FT REPEAT 534 557 2.
 SQ SEQUENCE 563 AA; 63929 MW; 866FE55704A1DE4B CRC64;
 Query Match 26.8%; Score 533; DB 1; Length 563;
 Best Local Similarity 31.0%; Pred. No. 1.1e-34;
 Matches 130; Conservative 58; Mismatches 158; Indels 74; Gaps 13;
 QY 1 AGGY-----W-HTSREILDANNVPRVIRAGINWFGFCNTCTVTVGLKSRDYRSM 50
 DB 27 AEGSYADLAEPDDDLHVEGTNIIVDKYGNKWTGANGWFGNCRERMLDLSVHSDIIADI 86
 QY 51 DQIKSLGYNTIRLPYSDDIL---KPGTMPNSINFRQWQDQLGLTSLQVMDKIVAYAGQI 107
 DB 87 ELVADKGINVVRPIATDLIYANSQGYTPSTDTSYNNPALAGNSVELFNFLNFKRV 146
 QY 108 GLRIILDRHPCGSGQS---ALWYTSVSSEATWISDLQALQRYKGNPTVVGPDLLNEPH 164
 DB 147 GIKVILDVHSPETDNOGHYPLWYNTTITEIFKAWVVAERYKNDTIIIGFDLKNESH 206
 QY 165 -----DPCWCGCGDPSIDWLAERAGNVLNVNENLLIFVEGVQSY----- 206

RESULT 8
 GUNC_PSEFL
 ID GUNC_PSEFL STANDARD; PRT; 748 AA.
 AC P27033;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase) (EGC) (Cellodextrinase C).
 GN CEUC.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.
 RC STRAIN=Sp. Cellulosa;
 RX MEDLINE=92061996; PubMed=1953673;
 RA Ferreira L.M.A., Hazlewood G.P., Barker P.J., Gilbert H.J.;
 RT "The cellodextrinase from Pseudomonas fluorescens subsp. cellulosa
 RT consists of multiple functional domains.";
 RL Biochem. J. 279:793-799 (1991).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
 CC domain.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61299; CAA43597.1; -;
 DR PIR; S19652; S19652.
 DR HSP; P07986; IEXG.
 DR InterPro; IPR001919; Bac_celose-bind.
 DR InterPro; IPR009031; CBDX.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR002883; Dockerin CBD 5.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00137; CBM_10; 1.
 DR Pfam; PF00553; CBM_2; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR SMART; SM00637; CBD II; 1.
 DR PROSITE; PS00561; CBD BACTERIAL; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 38
 FT CHAIN 39 748 ENDOGLUCANASE C.
 FT DOMAIN 39 136 CELLULOSE-BINDING.
 FT DOMAIN 137 179 SER-RICH (LINKER).
 FT DOMAIN 228 280 SER-RICH (LINKER).


```

CC 4 GYHTSGREILDANNVPRVIRAGINWFGFETCNVYVHGLWDRSRMLDQIKSIQYNTIRL 63
CC 37 GYHTSGREILDANNVPRVIRAGINWFGFETCNVYVHGLWDRSRMLDQIKSIQYNTIRL 96
CC 64 PYSDDLKPGMPNSINFRWNQDQGLTSLQVMDKIVAVAGQIGLRIILDRRDCSGQ 123
CC 97 PYSNQLFSSRRPSIDYHK-NPDLVGLNFIQMDKIEKAGQIGLRIILDRRDCSGQ 155
CC 124 SALWYTSVSSEATWISDQALQAKYKGNPTVGFDLNHPDPCWCGDPSIDWRLAAE 183
CC 156 SELWYTSQYPSRSWISDWKMDLADYKKNPTVIGADLNHPDPCWCGDPSIDWRLAAE 215
CC 184 RAGNAVLNPNLLIFVEGVQ--SYNGDSYVWGGNLOQAGQYVPLVNPRLVYSAHDY 240
CC 216 RAGNAVLNPNLLIFVEGVQ--SYNGDSYVWGGNLOQAGQYVPLVNPRLVYSAHDY 275
CC 241 ATSVYPTWFSDDPTFPNMPGINKWGYLNFQNIAPVWLGEFG--TLQSTTDTQWLKT 298
CC 276 GPGVSSQPFNDFPNSLPAIMQWGTWYISKQNIAPVWLGEFGNVDLSCPEGKQNA 335
CC 299 LVQYLRPTAQVAGSFWTWSNPDGSDGTGGILKDDQWQVTDVTKQGLYAPIKSSI 354
CC 336 LVHYI-----GANNLYFTYWSLNPNSGDTGGLLDDWTTWNRPKQDMLGRMKPV 385

RESULT 3
ID GUNA_XANCP STANDARD; PRT; 484 AA.
AC P19487;
DT 01-FEB-1991 (Rel. 17, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major extracellular endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-
DE beta-glucanase) (Cellulase).
GN ENKCA OR XCC3521.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-61.
RX MEDLINE=90323605; PubMed=2373365;
RT "Nucleotide sequence of the engXCA gene encoding the major
RT endoglucanase of Xanthomonas campestris pv. campestris."
RL Gene 89:53-59(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.F.,
RA Fortighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A.F., Sena T.A.D., Silva C., de Souza R.F.,
RA Spinoza L.A.F., Takita M.A., Tamara R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- MISCELLANEOUS: THE C-TERMINAL REGION OF THE PROTEIN IS NOT CRUCIAL
CC FOR ACTIVITY.

```

PRO/SER/THR-RICH (LINKER).
CELLULOSE-BINDING (BY SIMILARITY).
PROTON DONOR.
NUCLEOPHILE.

Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGYWHTSGREILDANNVVRVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
Db 42 AGGYWHTSGREILDANNVVRVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 101
QY 61 IRLPYSDIILKPGTMPNSINFRONQDLOGTSLQVMDKIVAYAGQIGLRIIILDRHRPDC 120
Db 102 IRLPYSDIILKPGTMPNSINFRONQDLOGTSLQVMDKIVAYAGQIGLRIIILDRHRPDC 161
QY 121 SGQSALWYTSVSSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPSPIDWRL 180
Db 162 SGQSALWYTSVSSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPSPIDWRL 221
QY 181 AAERAGNAVLSVNPENLLIFVEGVOSYNGDSYVWGGNLOAGAGQYPPVNLNPNRLVYSADY 240
Db 222 AAERAGNAVLSVNPENLLIFVEGVOSYNGDSYVWGGNLOAGAGQYPPVNLNPNRLVYSADY 281
QY 241 ATSVYPTWFSDDPTFPNNMPCGNKNGYLFNQNIAVWLGEGFTTQSTTQDQTLKTLV 300
Db 282 ATSVYPTWFSDDPTFPNNMPCGNKNGYLFNQNIAVWLGEGFTTQSTTQDQTLKTLV 341
QY 301 QYLRPTAQYGADSFQWTFWNNPDGSGTGILKDDQVTDVTKDGYLAPIKSSIFDPV 358
Db 342 QYLRPTAQYGADSFQWTFWNNPDGSGTGILKDDQVTDVTKDGYLAPIKSSIFDPV 399

RESULT 2

GUN_PABPO
ID_GUN_PABPO STANDARD; PRT; 397 AA.
AC P23548;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-PEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase (EC 3.2.1.4) (endo-1,4-beta-glucanase) (Cellulase).
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90170877; PubMed=2307659;
RA Baird S.D., Johnson D.A., Seligy V.L.;
RT "Molecular cloning, expression, and characterization of
endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
circulans";
RL J. Bacteriol. 172:1576-1586(1990).
CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -! SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
hydrolases).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M33791; AAA22631.1; --
DR PIR; A35136; A35136.
DR HSP2; P54583; LECE.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 194 194 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 317 317 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 397 AA; 44357 MW; B9C2E802C04FOA2A CRC64;

Query Match 52.8%; Score 1048; DB 1; Length 397;
Best Local Similarity 54.2%; Pred. No. 1.3e-75;
Matches 193; Conservative 45; Mismatches 106; Indels 12; Gaps 4;

401 461
462 562
203 323
75 161
209 212
48 50
51 52
53 55
57 58
61 62
65 68
72 72
73 73
75 76
80 84
87 96
97 98
101 107
108 111
113 114
127 129
130 131
134 147
148 149
151 158
160 160
161 162
163 163
175 188
189 191
193 194
195 199
206 207
209 209
215 217
218 218
219 230
231 233
235 236
242 242
244 244
246 247
248 249
250 251
254 255
257 257
259 263
274 279
283 285
289 292
294 299
300 307
308 308
309 313
314 315
319 323
331 343
344 344
347 350
351 352
356 359
367 369
373 373
375 376
380 380
382 387
388 388
389 391
SQ SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;
Query Match 99.6%; Score 1979; DB 1; Length 562;
Best Local Similarity 99.7%; Pred. No. 4.9e-149;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.33333 Seconds
(without alignments)
2236.934 Million cell updates/sec

Title: US-09-997-504A-14

Perfect score: 1986
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	562	1 GUN1_ACICE	P54583 acidothermu
2	1048	52.8	397	1 GUN1_PAEPO	P23548 paenibacill
3	803.5	40.5	484	1 GUN1_XANCP	P19487 xanthomonas
4	626.5	31.5	1039	1 GUN1_CALSA	P10474 c endogluca
5	608.5	30.6	565	1 GUN1_CLOTM	Q05332 clostridium
6	549.5	27.7	747	1 GUN1_CELFI	P50400 cellulomona
7	533	26.8	563	1 GUN1_CLOTM	P04956 clostridium
8	298.5	15.0	748	1 GUN1_PSEFL	P27033 pseudomonas
9	184	9.3	424	1 GUN1_RALSO	P58599 ralstonia s
10	166	8.4	426	1 GUN1_RALSO	P17974 ralstonia s
11	156.5	7.9	814	1 GUN1_CLOTM	P10477 clostridium
12	146.5	7.4	406	1 GUN1_RUMAL	P16216 ruminococcu
13	143.5	7.2	584	1 GUN1_CLOCE	P25472 clostridium
14	137	6.9	900	1 GUN1_CLOTM	P16218 clostridium
15	134	6.7	409	1 GUN1_RUMAL	P23661 ruminococcu
16	132	6.6	438	1 EXG_CANAL	P29717 candida alb
17	128.5	6.5	388	1 GUN1_HUMIN	Q12624 humicola in
18	128	6.4	448	1 GUN1_CLOSA	P15704 clostridium
19	124	6.2	440	1 GUN1_CLOCL	P28621 clostridium
20	123	6.2	825	1 GUN1_BACS4	P19570 bacillus sp
21	122.5	6.2	517	1 GUN1_CLOLO	P54937 clostridium
22	120	6.0	336	1 GUN1_RUMFL	P16189 ruminococcu
23	118	5.9	501	1 YBQ6_YEAST	P38081 saccharomyc
24	117	5.9	445	1 SPRI1_YEAST	P32603 saccharomyc
25	116	5.8	562	1 EXG2_YEAST	P52911 saccharomyc
26	114.5	5.8	658	1 GUN1_FIRSU	P14250 fibrobacter
27	112	5.6	764	1 VIS7_YEAST	P40566 saccharomyc
28	110.5	5.6	505	1 GUNV_ERWCA	Q47096 erwinia car
29	110.5	5.6	532	1 HEXA_DICDI	P13723 dictyostelli
30	109.5	5.5	418	1 GUN2_TRIRE	P07982 trichoderma
31	109.5	5.5	429	1 GUN1_BUTFI	P22541 butyrivibri
32	108.5	5.5	459	1 GUN1_STRLI	P27035 streptomyc
33	108.5	5.5	466	1 GUN5_THEFU	Q01786 thermomonos

RESULT 1

ID	GUN1_ACICE	STANDARD;	PRT;	562 AA.
AC	P54583;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1).			
OS	Acidothermus cellulolyticus.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Frankineae; Acidothermaceae; Acidothermus.			
OX	NCBI_TaxID=28049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 43068 / 11B;			
RA	Laymon R.A.; Himmel M.E.; Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=96346058; PubMed=8718854;			
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplus P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from			
RL	Acidothermus cellulolyticus in complex with cellobiose."			
CC	biochemistry 35:10648-10660(1996).			
CC	-!- FUNCTION: Has a very high specific activity on			
CC	carboxymethylcellulose.			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.			
CC	-!- MISCELLANEOUS: Thermostable enzyme with an optimal temperature of			
CC	81 degrees Celsius.			
CC	-!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl			
CC	hydrolases).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U33212; AAA75477.1; --			
DR	PDB; 1ECE; 14-OCT-96.			
DR	PDB; 1COD; 23-JUL-99.			
DR	InterPro; IPR001919; Bac_celose-bind.			
DR	InterPro; IPR008965; Cellul_bind.			
DR	InterPro; IPR001547; Glyco_hydro_5.			
DR	Pfam; PF00553; CBM 2; 1.			
DR	Pfam; PF00150; cellulase; 1.			
DR	SMART; SM00637; CBD_II; 1.			
DR	PROSITE; PS00559; GLYCOSYL_HYDROL_F5; 1.			
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	CATALYTIC.			

ALIGNMENTS

This Page Blank (uspto)

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neco, E.; Docena, C.; El-Borri, H.; Facincani, A.P.; Ferreira, A.J.S. Submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Pranca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigz Chado, M.A.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0810

Query Match 16.3%; Score 324; DB 2; Length 565;
Best Local Similarity 29.6%; Pred. No. 1.2e-17;
Matches 103; Conservative 54; Mismatches 135; Indels 56; Gaps 12;

Qy 12 EILDANNV-----PVRIAGINWGFETCNVYVHGLWSRDSYRMLDQIKSLGYNTIRL 63
Db 4 ETWDAQNAVSDTHFEVFLH--GVNWRGLETAQHLLOGLDQRPWREVLVDQWQSLGINAIRL 61

Qy 64 PYSDDILKPTMPNSINFRONODLOGLSLQWMDKIVAYAGQIGLRIILDRHPC-SG 122
Db 62 PLCSDTLH-GTMTPLNDLVR-NPDLKGRITALQIADAIIDEAGKGRMVLAYHVECPD 119

Qy 123 QSALWNTSSVSEATWISDQALAAQRYKGNPTVY-GFDLNEHPDPCACGPGSPIDWRLA 181
Db 120 GNPLRSVDESEHQWISDVQFIISHRAQKQVGVGVLDADMANRPFQSGDSTPDNVRV 179

Qy 182 AERAGNAVLSVNPNLIFVEGV-----QSYNGDSYVWGNLQAGQVPPVNLVNPNR-LVTS 236
Db 180 VERAAAILAAMTPEWLGVPGLNPPCLDASAPISDNIQ--SOHCVPLRIPARNLLM 237

Qy 237 AHVATSVYQTFWFSPTFFNNPFGINKWGLFNQNIAPVWLGFG--TTLQSTTDQ 293
Db 238 PHFAGTDIDTE-----AALGAFSGKQTVLPNSLDA 267

Qy 294 TWLTKTVQLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDQWQTVDT 341
Db 268 TDAEQLAHRIDALLAFGR--QGFYGSWMTSAQMPFGMLNDGRTPT 313

RESULT 13

S19652
C;Species: Pseudomonas fluorescens
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C;Accession: S19652; S18949
R;Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.
Biochem. J. 279, 793-799, 1991
A;Title: The cellobiohydrolase from Pseudomonas fluorescens subsp. cellulosa consists of
A;Reference number: S19652; MUID:92061996; PMID:1953673
A;Accession: S19652
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-748 <FER>
A;Cross-references: EMBL:X61299
R;Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.
submitted to the EMBL Data Library, August 1991
A;Description: The cellobiohydrolase from Pseudomonas fluorescens subsp cellulosa consist
A;Reference number: S18949
A;Accession: S18949
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262, 'C', 264-291, 'K', 293-748 <FER>
A;Cross-references: EMBL:X61299; NID:945501; PID:g45502
C;Genetics:
A;Gene: celC
C;Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain hom
F;38-134/Domain: Bacterial cellulose-binding domain homology <BC>

F;183-220/Domain: glycosidase GWGW domain homology <GWG>
F;39-133/Disulfide bonds: #status predicted

Query Match 15.0%; Score 298.5; DB 2; Length 748;
Best Local Similarity 23.4%; Pred. No. 1.8e-15;
Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;

Qy 7 HTSGREILDANNVPVRIAGINWGFETCN-----YVHGLW-----S 43
Db 290 NTQGNLTKDQGLLPARGC--NWFELEGRHEPSNDADNPGSGAPMELVAGNMWVNNQSGS 347

Qy 44 RYRSMLDQIKSLGYNTIRLPYSDIIL-----KGTMPNSINFRQNDLOGLSLQWMDK 99
Db 348 RTIOQTMTLQKQGITWRLPIAQTLDANDPQGRSENKHKHSIRO-----NARQALD 403

Qy 100 IVAYAGQIGLRIILDRH-----RPDCSGQSALWY-----TSSVS 133
Db 404 FIKLADQNDIQIFDIHSCSNYVGVWRAGRLDARPPYVDANRVGYDFTRREYSCSATNPS 463

Qy 134 EAT-----WISDQLALA--QRYKGNPTVYVGFDLNHPDPCACGPGSPIDWRLA 182
Db 464 SVTRIHAHYDKQKWLANLREIAGLSAKLVGNLIGIDVFNPEYD-YTWA-----EWKGV 516

Qy 183 ERAGNAVLSVNPNLIFVEGVQSYNGDSY-----WVWGNLQAGQY 223
Db 517 EEAQALINEVNPMLIIVEGI-SANANTQDGTPTSVVPHGSTDLNPNNGENLYEAGAN 575

Qy 224 PVNLNVP-NRLVYSAHDAVATSVYQTFWFSPT-----FPNNWPG 261
Db 576 PP-NIPDRLLFPSPHTYGPVSFVQRFMDPAQTECAGLEGEAAQARCRIVINPTVLEQ 633

Qy 262 IWNKNGVLFNQNIAPVWLGFG-----TTLQSTTDQTWLTKTVQL 303
Db 634 GWEHFGVLRGLVG--ILIEFGGNWDPGAKSSQADRNASHITTVNDQWQQAASYSF 692

Qy 304 RPTAQYAGDSFQWTFWSNPDSDGTGILKDDQWQTVDTKDGYLAPIKSSIFPDV 358
Db 693 K-----RKGINA---CVWSMNPESADTMG-----YLTTP-----WDPV 722

RESULT 14

E86265
protein F3f19.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E86265
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. xer, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: AB6141; MUID:21016719; PMID:11130712
A;Accession: E86265
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-522 <STO>
A;Cross-references: GB:AE005172; NID:g4850396; PIDN:AAD31066.1; GSPDB:GN00141
C;Genetics:
A;Gene: F3f19.15
A;Map position: 1

Query Match 13.0%; Score 257.5; DB 2; Length 522;
Best Local Similarity 25.1%; Pred. No. 1.9e-12;
Matches 94; Conservative 59; Mismatches 150; Indels 71; Gaps 18;

Qy 8 TSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDSYRMLDQIKSLGYNTIRLPYSD 67
Db 7 TSSRWIVDENGRLVKLVCAW--PSHLQPYVAEGLSKQPVDAVAKKIVEMGNCVRLTWPL 65

Db 343 YAQPFKEDFTWTLDDYDWTWYIHDQIAPILIGWGHMDGKNQKMTLLRDYI 402
QY 304 RPTAQYAGDSFOWTFSWNPDSGDTGGILKDDQWQVTDVTDGYLAP 349
Db 403 -----VQNRHHHTFCWCFNPNFSGDTGGILGNDWSTDEAKYALLKP 442

RESULT 10
B47093
cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Cellulomonas fimi
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
C:Accession: B47093
R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),
F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>
A:Reference number: A47093; MUID:93209933; PMID:8458833
A:Accession: B47093
A>Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-747 <MEI>
A:Experimental source: ATCC 484
A>Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIIP:128122)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>
F:645-744/Disulfide bonds: #status predicted

Query Match 27.7%; Score 549.5; DB 2; Length 747;
Best Local Similarity 34.5%; Pred. No. 3.6e-35;
Matches 130; Conservative 59; Mismatches 143; Indels 45; Gaps 14;
QY 1 AGGGYW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYN 59
Db 40 ATGDDNLHVEGNTIVDSTGCKEALISGVNWFNFASERVFHGLWSGNTITQIQQAQGIN 99
QY 60 TIRLPYSDDL---KPGTM--PNSINFRQMDQLGTSLOVMDKIVAYAGQIGLRIILD 114
Db 100 VVRVPSVQLLEWKAQTFKPNVNTY--ANPELEGKNSLQIFFEYWLTLCKQYGIKVFELD 157
QY 115 RH--RPDCSGQ-SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPH-----D 165
Db 158 VHSRADNSGHVYNNWKGDDITTEDVYEGWAAATRWKDDDTIVGADIKNEPHGTQGST 217
QY 166 PACWCGDPSIDWRLAERAGNAVLSVNPNNLLIFVGVQSYN-----GDSY--W 212
Db 218 RAKWDGTTDKNFHFAETAASKILLAINPNLVFVGVGVVPEWTSGLTDYGTW 277
QY 213 WGNLQAGAGQYPPVLV--PNRLVYSAHDYATSVYPTWF--SDPTFPNPNKNNKNGYL 270
Db 278 WGNLVRGRDHPIDIGAHQDQVYSPHDYGPVLDQKWFQKDFKASLTADVWGPNNLFI 337
QY 271 FNQNIAPVNLGFEFGTTL-QSTTDQTW---LKTLYQLRPTAQYAGDSFQWTFWSNPDSDG 326
Db 338 HDEDIAPLLIGEWGRLQDPRQDKWMAALDLVAER-----LSQTFWVLNENSG 388
QY 327 DTGGILKDDWQVTDVTK 343
Db 389 DTGGLLLDDWKTDEVK 405

RESULT 11
CZCLSM
cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase B precursor
C:Species: Clostridium thermocellum
C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 18-Jun-1999

C:Accession: A23512
R:Grepinet, O.; Beguin, P.
Nucleic Acids Res. 14, 1791-1799, 1986
A:Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endoglucan
A:Reference number: A23512; MUID:86148508; PMID:3453102
A:Accession: A23512
A:Molecule type: DNA
A:Residues: 1-563 <GRE>
A:Cross-references: GS:X03592; NID:940668; PIDN:CAA27266.1; PID:940669
C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase co
C:Genetics:
A:Gene: celB
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans
A:Pathway: cellulose degradation
C:Superfamily: cellulase B; Clostridium cellulase repeat homology
C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-563/Product: cellulase B #status predicted <MAT>
F:502-525/Domain: Clostridium cellulase repeat homology <CCR1>
F:534-557/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 26.8%; Score 533; DB 1; Length 563;
Best Local Similarity 31.0%; Pred. No. 4.9e-34;
Matches 130; Conservative 58; Mismatches 158; Indels 74; Gaps 13;
QY 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSML 50
Db 27 AEGSYADLAEPPDDWLVHVEGNTIVDKYGNKWTGANWFGNCRERMLDSTHSDIADI 86
QY 51 DQIKSLGYNTIRLPYSDDL---KPGTMPSNIFRQMDQLGTSLOVMDKIVAYAGQI 107
Db 87 ELVADKGINVWRMPIATDLIYAWSQGIYPPSTDTSTNNPALAGLSYELFNEMFNKRV 146
QY 108 GLRIILDRHRPCSGQS---ALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPH 164
Db 147 GIKVILDVHSPETDNGHNPYLPYANTITEEFKAWVWVAERYKNDTTFIIGFDLKNPH 206
QY 165 -----DPACWCGDPSIDWRLAERAGNAVLSVNPNNLLIFVGVQSY----- 206
Db 207 TWTGTMKIIQAQSAINDDSNHPNNKVAEETALALEVHPNVLIFVGVEMYPKDGIMDD 266
QY 207 -----NGDSY--WWGMLQAGQYPPVLV--VPNRLVYSAHDYATSVYPTWFSDP 253
Db 267 ETFTDPTSGTNNDDYGNWNGNLRGVYDFINLGRYQSOLVYSPHDYGFIVYEQDNFKGD 326
QY 254 TFPNN-----MPCGNKNGVILFNQNIAPVWLCEFGTTLQSTTDQTLWTLVQYLRP 305
Db 327 FITANDEQAKRIIYEOCWEDNWAYIMEEGISPLLGEWGMTEG--GHPLLDNLKYLRC 384
QY 306 TQAYGADS---FQWTFWSNPNPDGDTGGILKDD-----WQTVDTVKDGYLAPI 350
Db 385 MRDFILENKYKLHHTFCWCFNPNFSGDTGGILFTDEGTPPGGRDLKWN--DNKYDNYLYPV 442

RESULT 12
B82761
extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a50
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82761
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A:Title: the genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82761
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-565 <SIM>
A:Cross-references: GB:AE003920; GB:AE003849; NID:90105697; PIDN:AAF83620.1; GSPDB:GN0012
A:Experimental source: strain 9a50
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.5%; Score 626.5; DB 2; Length 915;
Best Local Similarity 35.5%; Pred. No. 4.3e-41;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLASRDYRSMLODKISLGYNTIRLP 64
DB 505 WLYVSGNKIVDKGRPWLTGYNWFGYNTGTVNFDGWSNCKLDTLAEIANRGFNLLRVP 564
QY 65 YSDIL---KPGTMPN-SINFRQMNQDLOGLSLQVMDKIVAYAGQIGLRIILDRH--RP 118
DB 565 ISAEILNWSQGIYPKENINY-VYNPELGKNSLEVDIVVQCKEVGLKIMDIHSIKT 623
QY 119 DCSGO-SALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDP-----ACWG 170
DB 624 DAMGHIYPVWYDEKFTPEFYKACEWITNRYKNDDTIIAFDLKNEPHGKPMQDTTFAXWD 683
QY 171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNLL 217
DB 684 NSTDIINWKYAAETCAKRIILNINPNLLIVIEGIEAYPKDDVTWTSKSSDDYISTWGGNLL 743
QY 218 QGAGQYPPVL-NVFNRLVYSAHDYATSVYPTQWFSPTFPNN--MPGIWKNWGYLFNQ 274
DB 744 RGVKYPINLKYQNVVSPHDYGPSVYQWPWFY-PGFTKESLLQDCWRPNWAYIMEEN 802
QY 275 IAPVWLGEFTTLOSTDTQWLKTLVOYLRLPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB 803 IAPLLIGEWGHLGDADNEKWKYLRDYII-----ENHIHTFWCFNANSDDTGLGVY 856
QY 335 DMQTVDTVKDGYLAP 349
DB 857 DFTTWDEKYSFLKP 871

RESULT 8

S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C:Species: Caldocellum saccharolyticum
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo-
A:Reference number: S02711; MUID:89098398; PMID:2789517
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1-1039 <SAU>
A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C:Genetics:
A:Gene: celB
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: Glycosidase; hydrolase; polysaccharide degradation
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1039/Product: cellulase #status predicted <MAT>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.5%; Score 626.5; DB 2; Length 1039;
Best Local Similarity 35.5%; Pred. No. 5.1e-41;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLASRDYRSMLODKISLGYNTIRLP 64
DB 629 WLYVSGNKIVDKGRPWLTGYNWFGYNTGTVNFDGWSNCKLDTLAEIANRGFNLLRVP 688
QY 65 YSDIL---KPGTMPN-SINFRQMNQDLOGLSLQVMDKIVAYAGQIGLRIILDRH--RP 118
DB 689 ISAEILNWSQGIYPKENINY-VYNPELGKNSLEVDIVVQCKEVGLKIMDIHSIKT 747

QY 119 DCSGO-SALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDP-----ACWG 170
DB 748 DAMGHIYPVWYDEKFTPEFYKACEWITNRYKNDDTIIAFDLKNEPHGKPMQDTTFAXWD 807
QY 171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNLL 217
DB 808 NSTDIINWKYAAETCAKRIILNINPNLLIVIEGIEAYPKDDVTWTSKSSDDYISTWGGNLL 867
QY 218 QGAGQYPPVL-NVFNRLVYSAHDYATSVYPTQWFSPTFPNN--MPGIWKNWGYLFNQ 274
DB 868 RGVKYPINLKYQNVVSPHDYGPSVYQWPWFY-PGFTKESLLQDCWRPNWAYIMEEN 926
QY 275 IAPVWLGEFTTLOSTDTQWLKTLVOYLRLPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB 927 IAPLLIGEWGHLGDADNEKWKYLRDYII-----ENHIHTFWCFNANSDDTGLGVY 980
QY 335 DMQTVDTVKDGYLAP 349
DB 981 DFTTWDEKYSFLKP 995

RESULT 9

A40589
cellulase (EC 3.2.1.4) - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium thermocellum
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jun-2000
C:Accession: A40589; S31361
R:Remaite, M.; Beguin, P.
J. Bacteriol. 175, 3353-3360, 1993
A:Title: Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri-
A:Reference number: A40589; MUID:93273701; PMID:8501039
A:Accession: A40589
A:Molecule type: DNA
A:Residues: 1-566 <LEM>
A:Cross-references: GB:X69390; NID:g40677; PIDN:CAA49187.1; PID:g40678
C:Genetics:
A:Gene: celG
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A:Pathway: cellulose degradation
C:Superfamily: cellulase B; Clostridium cellulase repeat homology
C:Keywords: Glycosidase; hydrolase; polysaccharide degradation
F:503-526/Domain: Clostridium cellulase repeat homology <CCR1>
F:536-559/Domain: Clostridium cellulase repeat homology <CCR>

Query Match 30.6%; Score 608.5; DB 2; Length 566;
Best Local Similarity 34.5%; Pred. No. 5.8e-40;
Matches 140; Conservative 54; Mismatches 143; Indels 69; Gaps 15;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLASRDYRSMLODKISLGYNTIRLP 64
DB 44 WLHCKGKIYDMYGNVWLTGANWFGNCFHGAW-YDVKTILTSLADRGINLLRIP 102
QY 65 YSDIL---KPG-----TMPNSINFRQMNQDLOGL-----LTSQVMDKIVAYAGQI 107
DB 103 ISTEILYSWMLGKNPVSSTASNNPYPHYVNPDPYPTDDVKNSEIFDIINGYCKEL 162
QY 108 GLRIILDRHPRDC--SGQS-ALWY---TSS---VSEATWISDLQALAQRYKGNPTVGF 158
DB 163 GIKWMDIHSPPDANNSGHNYELWYKSTGCTGVVTTKMWIDTLVWLADKYNKDDTIIAFD 222
QY 159 LHNEPHDP-----ACWGGCDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSY- 206
DB 223 LKNEPHGRKRYTAEPVKLLAKWDNSTDENNNKYAAETCAKALLENPKVLIVIEGEVQY 282
QY 207 -----NGD-----SYWGGNLLQAGQYPPVVLNVNPN--RLVYSADHYATSV 244
DB 283 KTEKGYTDTDPDWGATGDASPWYSAWGGNLRGVKDYPIDGLPLNSQIVVSPHDYGPSV 342
QY 245 YPQTFW-SDPTFPNMPGIWKNWGYLFNQIAPVWLGEFTTLOSTDTQWLKTLVQYL 303

Db 26 YSISHGKVIDDKGN-QIOLKGSWFGFETTHNVHGLWTNWKFEFTIQISQMGNAVRLP 84
 QY 65 YSDILLKPGTNPNSINFRQWQDLOGTSLQWMDKIVAYAGQIGLRILDRHPDSCGQS 124
 Db 85 FCPASLNSNTSPSSIDYNR-NPDYQGLSSQLQIMDKVVKELSDRGSIYVIMDHPDCAAIS 143
 QY 125 ALWYTSVSEATWISDLQALAOAKYKGNPTVVGDFLHNEPHDPAWCGGCDPSIDWRLAAER 184
 Db 144 ELWYTSVSEATWISDLQALAOAKYKGNPTVVGDFLHNEPHDPAWCGGCDPSIDWRLAAER 203
 QY 185 AGNAVLSPNPLIFVSGVSYNGDS-----YWGNGNLQAGQVYVNLNVP-NRLVYSAMD 239
 Db 204 AAAAILEAAPKWLIGVEGIGENPSCSSTIGHFWGENLEPMDCPT-LKVPADHLLLIPIHV 261
 QY 240 YATSVYPTWFSDETFNNMPEGIWKKWGYLFNQNIAPVWLGEFTTL--QSTTDQTLWK 297
 Db 262 YGPVYVQVYFNSPDPFNNMAAIWKHGFPAKAGYA-MAIGFPGKYGEGDPRDVAQN 320
 QY 298 TLVQLRPTAQY-ADSFQWTFWSPNPDSDGTGGLXDDMQTV 339
 Db 321 ALVDYL---ISIGVTDTF---YMSINASSDGTGLPRDDNNHV 357
 RESULT 5
 E75142
 endoglucanase PAB0632 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: E75142
 R:Anonymous, Genoscope
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A:Reference number: A75001
 A:Accession: E75142
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-514 <KAW>
 A:Cross-references: GB:A0248285; GB:A096836; NID:g5458067; PIDN:CAB49854.1; PID:g545836
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: celB-like; PAB0632
 C:Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 34.7%; Score 688.5; DB 2; Length 514;
 Best Local Similarity 41.6%; Pred. No. 2.7e-46;
 Matches 148; Conservative 58; Mismatches 121; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETTCNYYVHGLWSDRYSMGLDQIKSLGYNT 60
 Db 30 YTAENGILFVQNTTGEKKPLYLHGVSFGEPLKDHVYGLDKENWKDILKDKVRLGFNA 89
 QY 61 IRLPYSDILLKPGTNPNSINFRQWQDLOGTSLQWMDKIVAYAGQIGLRILDRHPD 118
 Db 90 IRLPFCSESIRPDRPSERINY-ELNPDKLNTLSLEIMEKIIEYANSIGLYILLDYHRI 148
 QY 119 DCSSQALWYTSVSEATWISDLQALAOAKYKGNPTVVGDFLHNEPHDPAWCGGCDPSIDW 178
 Db 149 GCEIEPLWYNTVSEQYIKDWIFLAKRPGKYPNVIGAIKKEPHGEAGWGTGDER-DF 207
 QY 179 RLAAERAGNAVLSPNPLIFVSGVSYNGDSY-----SYNGDSY--WGNLQAGQVYVNL 227
 Db 208 RLPAEKVGRILKVAPEHLFVEGTQYTHVFNIDEIEKKGWTFWGENLMGVKDYVPL 267
 QY 228 NVPNRLVYSAHDYATSVYPTWFSDETFNNMPEGIWKKWGYLFNQNIAPVWLGEFTTL 287
 Db 268 -PRGVVYSPHYGVSVMYMDYKSPDPFNNMPIIWEHFGYLTDLNLTLLV-IGEMGNGY 325
 QY 288 QSTTDQTLWKTLVQLRPTAQYADSFQWTFWSPNPDSDGTGGLXDDMQTVDTVK 343
 Db 326 EG-LDKVQDAFVKWLKKIY-----NFFVWCLNPSGDTGGIFLDDWKTAVNEK 375

RESULT 6

E97012
 probable non-processive endoglucanase family 5, secreted, Cella homolog secreted, docked in
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: E97012
 R:Nolling, J.; Bretton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: E97012
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-482 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78888.1; PID:g15023812; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0912

Query Match 34.3%; Score 681; DB 2; Length 482;
 Best Local Similarity 37.6%; Pred. No. 9.5e-46;
 Matches 140; Conservative 71; Mismatches 127; Indels 34; Gaps 9;

QY 5 YHTSGREILDANNV-----PVRIAGINWFGFETTCNYYVHGLWSDRYSMGLDQIKSLGYNTIPLP 64
 Db 24 YLHSDGSKLLDDYGNQVRMTGIAWFGLETPNYCFHGLWANRLDNLINIVAGNFTLRVP 83
 QY 65 YSDILL---KPGT--MPNSINFRQWQDLOGTSLQWMDKIVAYAGQIGLRILDRHPD 119
 Db 84 LSVELVQWVRQGVYPTDPSIN-DVISPELKGNSQLLDDVIAYSKKVGVKVMLDHRIE 142
 QY 120 CSQALWYTSVSEATWISDLQALAOAKYKGNPTVVGDFLHNEPHDPAWCGG 172
 Db 143 SGGQTATWYTSKYTTDDYKQWYLDYKNDTVAIAIDFNEPHGKAYRAETSAKWN 202
 QY 173 DPSIDWRLAAERAGNAVLSPNPLIFVSGVSY-----NGDSY---WNGNLQAG 219
 Db 203 TDEDNRVYAEKVGKILIDNPMLIVGEVETPEKGTAGSINPDYGGWNGNLRG 262
 QY 220 AGQYPVVL-NVENRLVYSAHDYATSVYPTWTF-SOPTFNNMPEGIWKKWGYLFNQNIAP 277
 Db 263 VKDYPVDLAPYKYNQVYSPHDYGPVSDQVTFDGFTEQSLINDIWRPSWFIQEKNIAP 322
 QY 278 VMLGEGITLQSTTDQTLWKTLVQLRPTAQYADSFQWTFWSPNPDSDGTGGLXDDMQ 337
 Db 323 LLIGSWGNGMDGKGEQWTDMAKLI-----SDNRMHTFWCLNANSGLTGGILEYDFK 376
 QY 338 TVDTVTKDGYLAP 349
 Db 377 TIDTKLALVQP 388

RESULT 7
 A43802
 cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Caldocellum saccharolyticum
 C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
 C:Accession: A43802
 R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 56, 3117-3124, 1990
 A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "C"
 A:Reference number: A43802; MUID:91136262; PMID:2126700
 A:Accession: A43802
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translat
 A:Molecule type: DNA
 A:Residues: 1-915 <SAU>
 A:Cross-references: EMBL:X13602
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
 A:Pathway: cellulose degradation
 C:Superfamily: Streptomyces endo-1,4-beta-xylosidase A homology
 C:Keywords: Glycosidase; hydrolase; polysaccharide degradation

JH0158
cellulase (EC 3.2.1.4) precursor - Xanthomonas campestris pv. campestris
N;Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase
C;Species: Xanthomonas campestris pv. campestris
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C;Accession: JH0158
R;Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.
Gene 89, 53-59, 1990
A;Title: Nucleotide sequence of the engXCA gene encoding the major endoglucanase of Xanthomonas campestris pv. campestris
A;Reference number: JH0158; MUID:90323605; PMID:2373365
A;Accession: JH0158
A;Molecule type: DNA
A;Residues: 1-493 <G>
A;Cross-references: GB:M32700; NID:g155397; PIDN:AAA27612.1; PID:g155398
R;1-25/Domain: signal sequence #status predicted <SIG>
F;26-474/Product: cellulase #status predicted <MAT>
F;375-399/Region: proline/threonine-rich
C;Genetics:
A;Gene: engXCA
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A;Pathway: cellulose degradation
C;Superfamily: Xanthomonas campestris cellulase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-474/Product: cellulase #status predicted <MAT>
F;375-399/Region: proline/threonine-rich
Query Match 40.5%; Score 803.5; DB 1; Length 493;
Best Local Similarity 46.3%; Pred. No. 2.4e-55;
Matches 161; Conservative 61; Mismatches 102; Indels 19; Gaps 9;
QY 5 YWHTSGREILDANNVPRVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNTRILP 64
DB 26 YSINNSRQIVDDSGKVQLKGVNVFGFETCNVYVHGLWSDYRSMDDQIKSLGYNTRILP 85
QY 65 YSDDILKPGMPSINFRQNWQDLQGLTSQVMDKIVAYAGQIGLRIILDRHRPCSGOS 124
DB 86 FCPATLRSDTPASIDYR-NADLQGLTSQVMDKIVAYAGQIGLRIILDRHRPCSGOS 144
QY 125 ALMYTSVSATWISDLQALQARYKGNPTVVGFDLNEPHDPACWCGDPSIDWRLAER 184
DB 145 ELMYTSGYTEAQLADLRFLVANYKKNPVYVGLDLKNEPHDPACWCGDPSIDWRLAER 204
QY 185 AGNAVLSVNPILLIFEGVQ-----SYNGDSYVWGNLQAGQYVPLVNP-NRLVYSAH 238
DB 205 GSAALVAVAPKWLIAVEGITDNVPCVTNG-GIFWGNLQPLACTP--LIPANRLSLAHP 261
QY 239 DYATSVVPQWFSDFPFPNNMGIWKNWGLYFNQNIAPVWLGFEFTTL--QSTTDQTLW 296
DB 262 VYGPDIYVQSYFNSDFNPNMPLWESHFQGFAGTH--ALLGFEGKYGEGDARDKTWQ 319
QY 297 KTLVQYLRTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTV 339
DB 320 DALVKYLR---SKGIN--QGFYWSWNPNSGDTGGLRDDWTSV 357
RESULT 3
E71059
probable endo-1,4-beta-glucanase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: E71059
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71059
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-458 <KAW>
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30271.1; PID:g3257588
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:

A;Gene: PH1171

C;Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 38.4%; Score 762; DB 2; Length 458;

Best Local Similarity 45.0%; Pred. No. 3.9e-52;

Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPRVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNTRILPYS 67

DB 54 TSGEE-----TPHLFGVNWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNTRILPYS 107

QY 68 DILKPGTNPNSINFRQNWQDLQGLTSQVMDKIVAYAGQIGLRIILDRHRPCSGOSQALW 127

DB 108 ESVKFGTPIGIDYSK-NPDRLGLDSQIMEKIILKAGDLGIFVLLDYHIGCTHIEPLW 166

QY 128 YTSSVSEATWISDLQALQARYKGNPTVVGFDLNEPHD-----PACWCGDPS 175

DB 167 YTEDFSSEDFINTWIEVAKRFGKYNVIGADLRKNEPHSVTSPPAAYTGTGATWGMGNA 226

QY 176 IDWRLAERAGNAVLSVNPILLIFEGVQSYN-----GDSYVWGNLQAGQYVPLV 226

DB 227 TDWLAERIGKALKVAPHLVIFVEGTQTNPKTDSYKYNWGNLMAVKDIPV- 285

QY 227 LNVP-NRLVYSAHYATSVTPQWFSPT-FPNNMGIWKNWGLYFNQNIAPVWLGFE 284

DB 286 -NLPRNKLVSPHYGPDVYNQPYFGPAKGFDPDLPIWYHFGYKLELGYSVVIGFE 344

QY 285 TTL---QSTTDQTLWKLTVQVLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTV 339

DB 345 KGYHGDPDVIVQNKLVDMW--LENKFCDF--YASWNPDSGDTGGLQDDWTII 397

RESULT 4

E82759

endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: E82759

R;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82759

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-592 <SIM>

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincan, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaku

M.; Tshako, M.H.; Vallada, F.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0818

Query Match 35.0%; Score 695.5; DB 2; Length 592;

Best Local Similarity 42.6%; Pred. No. 9.2e-47;

Matches 146; Conservative 52; Mismatches 126; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPRVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNTRILP 64

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds
(without alignments)
2792.154 Million cell updates/sec

Title: US-09-997-504A-14
Perfect score: 1986
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGLAPIKSSIFDFV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1050	52.9	397	2 A35136	cellulase (EC 3.2.1.4)
2	803.5	40.5	493	1 JH0158	cellulase (EC 3.2.1.4)
3	762	38.4	458	2 E71059	probable endo-1,4-beta-glucanase
4	695.5	35.0	592	2 E82759	endo-1,4-beta-glucanase PAB0
5	688.5	34.7	514	2 E75142	probable non-proce
6	681	34.3	482	2 E97012	cellulase (EC 3.2.1.4)
7	626.5	31.5	915	2 A43802	cellulase (EC 3.2.1.4)
8	626.5	31.5	1039	2 S02711	cellulase (EC 3.2.1.4)
9	608.5	30.6	566	2 A40589	cellulase (EC 3.2.1.4)
10	549.5	27.7	747	2 B47093	cellulase (EC 3.2.1.4)
11	533	26.8	563	1 C2CLBM	extracellular endo
12	324	16.3	565	2 B82761	cellodextrinase C
13	298.5	15.0	748	2 S19652	protein F319.15 [
14	257.5	13.0	522	2 E86255	hypothetical prote
15	210	10.6	526	2 T51476	hypothetical prote
16	206.5	10.4	488	2 T51502	cellulase (EC 3.2.1.4)
17	166	8.4	426	2 A42649	cellulase (EC 3.2.1.4)
18	156.5	7.9	814	1 C2CLBM	mannan endo-1,4-be
19	154.5	7.8	516	2 JE0134	endo-1,4-beta gluc
20	154.5	7.8	722	2 H96986	endo-1,4-beta gluc
21	151.5	7.6	356	2 G82523	endoglucanase fami
22	147	7.4	1012	2 B37326	endoglucanase - Th
23	146.5	7.4	329	2 C72216	cellulase (EC 3.2.1.4)
24	146.5	7.4	406	2 A43732	cellulase (EC 3.2.1.4)
25	143.5	7.2	584	2 J01229	cellulase (EC 3.2.1.4)
26	139.5	7.0	669	2 D72278	endo-1,4-beta-mann
27	137	6.9	900	2 JH0157	cellulase (EC 3.2.1.4)
28	136.5	6.9	500	2 S22458	cellulase (EC 3.2.1.4)
29	134	6.7	370	2 G97001	endoglucanase fami

ALIGNMENTS

RESULT 1

A35136
cellulase (EC 3.2.1.4) - Bacillus polymyxa
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus polymyxa
C/Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 24-Sep-1998
C/Accession: A35136
R/Baird, S.D.; Johnson, D.A.; Seligy, V.L.
J. Bacteriol. 172, 1576-1586, 1990
A/Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase
A/Reference number: A35136; MUID:90170877; PMID:2307659
A/Accession: A35136
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-397 <BAI>
A/Cross-references: GB:M33791; GB:M33840
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A/Pathway: cellulose degradation
C/Superfamily: Xanthomonas campestris cellulase
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

30	134	6.7	409	2 S12018	endoglucanase B -
31	129.5	6.5	388	1 JC5461	cellulase (EC 3.2.1.4)
32	128.5	6.5	388	1 S43920	cellulase (EC 3.2.1.4)
33	128.5	6.5	455	2 S16559	cellulase (EC 3.2.1.4)
34	128	6.4	438	2 A47702	glucan 1,3-beta-gl
35	128	6.4	438	2 T52149	beta-glucanase [im
36	128	6.4	448	2 A27631	cellulase (EC 3.2.1.4)
37	126	6.3	656	2 AB1843	hypothetical prote
38	123	6.2	441	2 A44815	cellulase (EC 3.2.1.4)
39	123	6.2	825	2 JS0174	cellulase (EC 3.2.1.4)
40	122.5	6.2	517	2 I40798	hypothetical prote
41	120	6.0	411	2 D86153	endoglycosylcerami
42	119	6.0	482	2 JC7332	probable glucan 1,
43	118	5.9	501	2 S45914	exo-1,3-beta-gluc
44	117	5.9	445	2 A40639	glucan 1,3-beta-gl
45	116	5.8	562	2 S55516	glucan 1,3-beta-gl

Query Match 52.9%; Score 1050; DB 2; Length 397;

Best Local Similarity 54.2%; Pred. No. 7.9e-75; Indels 12; Gaps 4;
Matches 193; Conservative 46; Mismatches 105;

QY	4	GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLEDOIKSLGYNTIRL	63
DB	37	GYHTQGNKIVDESCKEAFGLNWFGLTPTNYTLHGLWSRMDMLDQVKKEGYNLRL	96
QY	64	PYSDDLKPGTNPNSINFRMNQDLQGLTSLOVMDKIYAYAGQIGRLILDRHRDCSQ	123
DB	97	PYSNQLFSSSPDSIDYHK-NPDLVGLNPQIMDKLEKAGQIGIILDRHRPGSGQ	155
QY	124	SALWTSVSSEATWISDLQALAQRYKGNPTVVFGLHNEPHDPACGCGDPSIDWRLAE	183
DB	156	SELWTSQYPSERWISDWKMLADRYKNPTVIGADLHNEPHGQASWGNGNASTDWRLAAQ	215
QY	184	RAGNAVLSNPMLLIFVSGVQ---SYNGDSYWGNGLOAGQYPVVLPVRLVYSAHDY	240
DB	216	RAGNAVLSNPMLLIFVSGVQ---SYNGDSYWGNGLOAGQYPVVLPVRLVYSAHDY	275
QY	241	ATSVTPQTFESPTPTNNMGGIWNKNWGLFNQNIAPVNLGEGF--TTLQSTDTDTLTKT	298
DB	276	GGCVSSQPFNWDPAFNSLPALWDQTWGYISKQNIAPVLVGEFGGNDVLSPEKQWNA	335
QY	299	LVQYLRLPTAQYAGDSFQWTFWSNPDGSGTGILKDDQWQTVDTVKDGLAPIKSI	354
DB	336	LVHYI-----GANNLYFTYSLNPNDSGDTGGLLLDDWTWNRPKQDMLGRIMKPV	385

RESULT 2

This Page Blank (uspto)

[illegible]

RESULT 13
 US-10-369-493-21618
 : Sequence 21618, Application US/10369493
 : Publication No. US20030233675A1
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yongwei
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Siater, Steven C.
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Chen, Xianfeng
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 : FILE REFERENCE: 38-10(52052)B
 : CURRENT APPLICATION NUMBER: US/10/369,493
 : CURRENT FILING DATE: 2003-02-28
 : PRIOR APPLICATION NUMBER: US 60/360,039
 : PRIOR FILING DATE: 2002-02-21
 : NUMBER OF SEQ ID NOS: 47374
 : SEQ ID NO 21618
 : LENGTH: 514
 : TYPE: PRT
 : ORGANISM: Pyrococcus abyssi
 US-10-369-493-21618

Query Match	34.7%	Score 688.5;	DB 15;	Length 514;
Best Local Similarity	41.6%;	Pred. No. 8.1e-59;		
Matches 148;	Conservative 58;	Mismatches 121;	Indels 29;	Gaps 10;
Qy	7	HTSGREILDANNV-----PVRIAGINWFGTETCYNVVHGLWSRDYRSMLDQIKSLG	VNT 60	
Db	30	YTAENGIIIVQNVITTEKKPIYLHGVSFGEFLKHVVYGLDKENWKDILKDVKRLG	ENA 89	
Qy	61	IRLPYSDDILKPGWPN--SINFQMNODLOGLTSLQWMDKIYVAGQIGRIILDRHP	118	
Db	90	IRLPFCSESIREDTRPSPERINY-ELNPDKNLTSLTEIMEKIIIFYANSIGHYILLDYH	148	
Qy	119	DCSGQSALWYTSVSEATWISDLOALAQYKGNPTVVGFDLHNEPHOPACWCGCDPS	IDW 178	
Db	149	GCSEIEPLWYTYENSEEQYIKDWIFLAKRFGKYPNVIGADIKNEPHGAGTGDGR-D	F 207	
Qy	179	RLAERAGNAVLSVNPNNLLIFVEGVQ-----SYNGDSYW--WGNLQACQYPVVL	227	
Db	208	RLFAEKVGRILKVAHPHLLIFVEGQYTHVPNIDIEIKKGWTFWNGENLWGVKDY	PVL 267	
Qy	228	NVPNRLVYSAHDYATSVPTQWFSFDPTEPNNNPGLVWKNQGYLFNQNLAPVWLGE	FGTTL 287	

```
Db      268 -PRGKVVSHPVYGSVTMMDFKSPDPFNMPPIWETHFGYLTDLNYTLV-IGEWGGNY 325  
        :|||:: ||| ::|||:: |||:: |||:: |||:: |||:: |||::  
  
Qy      288 QSTDTQTWLKTIVQLRYRETAQYGADSFQWTFSWNPDSGTGIGLKDDWGTVDTVK 343  
        :   |       :|||:: |||:: |||:: |||:: |||:  
  
Dd      326 EG-LDKVMQDAFVKVLKKIY----NFFVCWCLNPESGDGTGGIFLDMDKITVNWEK 375  
        :|||||:: |||:: |||:: |||:: |||:: |||:: |||:
```

```
RESULT 14  
US-10-369-493-9473  
; Sequence 9473, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTIVITY  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(520S2) B  
; CURRENT APPLICATION NUMBER: US /10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 9473  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-9473
```

Query Match	33.6%;	Score	667.5;	DB	15;	Length	369;
Best Local Similarity	40.5%;	Pred. No.	6e-57;				
Matches	143;	Conservative	58;	Mismatches	123;	Indels	29;
Gaps	12;						
Qy	5	YWHTSGREILDANNVPVRIAGINWFGP	ETCNYVVHGLWSRDYRSMLOIKSLGIVNTIRLP	64			
Db	26	YSISHGKVVDDKGN-QIOLRGVNWFGP	ETGHDVVVGLWARNWKKEFIITQLQGNGFNAIRLP	84			
Qy	65	YSDDILKPGTWPNSINFRQWNCIDLO	QSTLSQVMDKIVAYAGQIGILRIILDRHPRDCSQS	124			
Db	85	FCPANLNSNTSPSSIDYSR-NPDIQL	GLSSQLLDKVVKELSDRRMYVLDDHRRSDCSAIS	143			
Qy	125	ALWYTSSVSEATWTSIDLOALIAQR	YKGNPTVVFGDLHNEPHDPACWGCDDPSIDNRLAER	184			
Db	144	ELWYTDVYGEKQWIDDLRFVAHRY	TNVPGVIGLDVKNEPHGREATWGTGDPXTDMNTVAEH	203			
Qy	185	AGNAVLVNPENLLIFVEGVQSYNGDS	-----YWGGNLQAGQYVPWLNVN- NRLVY	235			
Db	204	AAAAIILEAAPKWLITGVSGI----	GENPICSTIGHFWGLENLEFMDCTP-- LKVPANHLUL	257			
Qy	236	SAHDYATSVYEPQTWFSDPPTFN	NNMPGWNKNWGLFNQNIAPVWLBEFTTL-- -QSTTDQ	293			
Db	258	MPHYVGPDIYVQYPNSPDPFN	NNAAIWDQHFGRFAKAGYA- MAIGREGGKYGEGDPDRI	316			
Qy	294	TWLKTLVOYLBPRTAQYG-ASQ	FQMTWFSWNPDSGDTGSILKDDQTVV-- -DTVK	343			
Db	317	AWONAFVDYL---ISIGVTDFAF---	YVAAANQNSGDTGVMGNDMTTPRDDVKV	363			

RESULT 15
US-10-369-493-9262
; Sequence 9262, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

Db 167 YTEDFSEEDFINTWIEAKRFKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLSNPDLIFVEGVQSYN-----GDSYWGGLQAGQYFV 226
Db 227 TDWNLAAERIGKALIKVAPHWLIIEVEGTQFTNPKTDSYKNGYNAMWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFSDEPT-FPNMMPGIWKNWGLYFNQNIAPVWLGEFG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGAKGPDNLDPIWYHHFGYVKLELGSYVVIIEFG 344
Qy 285 TTL---QSTTDQTKLVQYLRTAQYGADSFQWTFWSNPDSDGTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWONKLVDMW--IENKFCDF--YMSWNPDSGDTGGILQDDWTTI 397

RESULT 10

US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 38.4%; Score 762; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 3.9e-66;
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDRYSMLDQIKSLGYNTIRLPYS 67
Db 54 TSCEE-----TPHILFGVNWFGFETPNHVHGLWKRWNEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFRQMNQDLQGLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGQSA 127
Db 108 ESKVPGTQPIGIDYSK-NPDLRGLDSLQIMEKIIKAGDLGIFVLLDYHRIGCTHIEPLW 166
Qy 128 YTSSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEAKRFKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLSNPDLIFVEGVQSYN-----GDSYWGGLQAGQYFV 226
Db 227 TDWNLAAERIGKALIKVAPHWLIIEVEGTQFTNPKTDSYKNGYNAMWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFSDEPT-FPNMMPGIWKNWGLYFNQNIAPVWLGEFG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGAKGPDNLDPIWYHHFGYVKLELGSYVVIIEFG 344
Qy 285 TTL---QSTTDQTKLVQYLRTAQYGADSFQWTFWSNPDSDGTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWONKLVDMW--IENKFCDF--YMSWNPDSGDTGGILQDDWTTI 397

RESULT 11

US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20347
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20347

Query Match 38.4%; Score 762; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 3.9e-66;
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDRYSMLDQIKSLGYNTIRLPYS 67
Db 54 TSCEE-----TPHILFGVNWFGFETPNHVHGLWKRWNEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFRQMNQDLQGLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGQSA 127
Db 108 ESKVPGTQPIGIDYSK-NPDLRGLDSLQIMEKIIKAGDLGIFVLLDYHRIGCTHIEPLW 166
Qy 128 YTSSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEAKRFKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLSNPDLIFVEGVQSYN-----GDSYWGGLQAGQYFV 226
Db 227 TDWNLAAERIGKALIKVAPHWLIIEVEGTQFTNPKTDSYKNGYNAMWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFSDEPT-FPNMMPGIWKNWGLYFNQNIAPVWLGEFG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGAKGPDNLDPIWYHHFGYVKLELGSYVVIIEFG 344
Qy 285 TTL---QSTTDQTKLVQYLRTAQYGADSFQWTFWSNPDSDGTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWONKLVDMW--IENKFCDF--YMSWNPDSGDTGGILQDDWTTI 397

RESULT 12

US-10-369-493-17534
; Sequence 17534, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17534
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-17534

Query Match 35.0%; Score 695.5; DB 15; Length 592;

QY 181 AERAGNAVLSVNPGLLIFVEGVSQNGSDSYWGGNLOAGAGQYPPVNLVNPRLVYSADHY 240
DB 222 AERAGNAVLSVNPGLLIFVEGVSQNGSDSYWGGNLOAGAGQYPPVNLVNPRLVYSADHY 281
QY 241 ATSVYPTQWFSPTFPNNPGIWNKQWGLFNQNIAPVWLBFGTTLQSTTDTQWLTIV 300
DB 282 ATSVYPTQWFSPTFPNNPGIWNKQWGLFNQNIAPVWLBFGTTLQSTTDTQWLTIV 341
QY 301 QYLRPTAQYGAQSFQWTFWSPNPDSDGTGGILKDDQWQVTDVTKDGYLAPIKSIIDPV 358
DB 342 QYLRPTAQYGAQSFQWTFWSPNPDSDGTGGILKDDQWQVTDVTKDGYLAPIKSIIDPV 399

RESULT 2

US-10-360-101-212
; Sequence 212, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 212
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of cellulase
US-10-360-101-212

Query Match 52.8%; Score 1048; DB 15; Length 397;
Best Local Similarity 54.2%; Pred. No. 2e-94;
Matches 193; Conservative 45; Mismatches 106; Indels 12; Gaps 4;

QY 4 GYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRYSMLDQIKSLGYNITRL 63
DB 37 GYHTQGNKIVDESGKEAENGLNWFGLTPTNYTLHGLWSRMDMDLQVXKGYNLRL 96
QY 64 PYSDDLKPGTNPNSINFRQWQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDCSGQ 123
DB 97 PYSNQLFSDSSRPDSIDYHK-NPDLVGLNPQIMDKLIEKAGQGIQILDRHRPDCSGQ 155
QY 124 SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
DB 156 SELWYTSQYDESRWISDWKMLADRYKKNPTVIGADLHNEPHGQASWGTGNASTDMLAAQ 215
QY 184 RAGNAVLSVNPGLLIFVEGVQ---SYNGDSYWWGGNLOAGAGQYPPVNLVNPRLVYSADHY 240
DB 216 RAGNAVLSVNPGLLIFVEGVQDHVYVGNQNSQYWWGGNLTGVANYPVVLVDVFNRYVSPHY 275
QY 241 ATSVYPTQWFSPTFPNNPGIWNKQWGLFNQNIAPVWLBFGT---TTLQSTTDTQWLT 298
DB 276 GPGVSSQFWNDPAPNLPALINDQWGYISKQNIAPVLBFGGGRNVDSLSCPEGKQNA 335
QY 299 LVQYLRPTAQYGAQSFQWTFWSPNPDSDGTGGILKDDQWQVTDVTKDGYLAPIKSI 354
DB 336 LVHYI-----GANNLYFTYWSLNFNSGDTGGILLDDWTTWNRPKQDMLGRIMKPV 385

RESULT 3

US-10-369-493-16379
; Sequence 16379, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16379
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16379

Query Match 40.5%; Score 803.5; DB 15; Length 493;
Best Local Similarity 46.9%; Pred. No. 3.5e-70;
Matches 161; Conservative 61; Mismatches 102; Indels 19; Gaps 9;
QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRYSMLDQIKSLGYNITRLP 64
DB 26 YSINNSRQIVDDSGKVQLKGVNVFPGTGNHVMHGLWARNKDMIVQMGLGFNAVRLP 85
QY 65 YSDDLKPGTNPNSINFRQWQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDCSGQS 124
DB 86 FCPATLRSDTNPASIDYSR-NADLQGLTSLQILDKVIBEFNARGMYVLLDHTDTCAGIS 144
QY 125 ALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
DB 145 ELWYTSVTEAQWLAIDLRFVANRYKVPVYVGLDLKNEPHGATWGTGNAATDNKKAER 204
QY 185 AGNAVLSVNPGLLIFVEGVQ---SYNGDSYWWGGNLOAGAGQYPPVNLVNP-NELVYSAH 238
DB 205 GSAAVLAVAPKWLIAVEGITDNPVCSTNG-GIFWGGNLOPLACTP--LNPANLLIAPH 261
QY 239 DYATSVYPTQWFSPTFPNNPGIWNKQWGLFNQNIAPVWLBFGTTL--QSTTDTQWL 296
DB 262 VYGPVDFVQSVFNDSPNPNPAPWHERFGQFAGTH--ALLGFGGKYGGEGDARDKTWQ 319
QY 297 KTLVQYLRPTAQYGAQSFQWTFWSPNPDSDGTGGILKDDQWTV 339
DB 320 DALVKYLR---SKGIN--QGFYWSNPNPNSGDTGGILRDDWTSV 357

RESULT 4

US-10-369-493-16123
; Sequence 16123, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16123
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16123

Query Match 40.1%; Score 736.5; DB 15; Length 483;
Best Local Similarity 46.6%; Pred. No. 1.7e-69;
Matches 160; Conservative 61; Mismatches 103; Indels 19; Gaps 9;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds
(without alignments)
2716.899 Million cell updates/sec

Title: US-09-997-504A-14

Perfect score: 1986
Sequence: 1 AGGGWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	562	9	US-09-981-900B-5
2	1048	52.8	397	15	US-10-360-101-212
3	803.5	40.5	493	15	US-10-369-493-16379
4	796.5	40.1	483	15	US-10-369-493-16123
5	796.5	40.1	501	15	US-10-369-493-15740
6	796.5	40.1	518	15	US-10-369-493-15373
7	796.5	40.1	535	10	US-09-927-827-50
8	770	38.8	453	9	US-09-888-224-2
9	762	38.4	458	14	US-10-293-344A-2
10	762	38.4	458	15	US-10-369-493-1281
11	762	38.4	458	15	US-10-369-493-20347
12	695.5	35.0	592	15	US-10-369-493-17534
13	688.5	34.7	514	15	US-10-369-493-21618
14	667.5	33.6	369	15	US-10-369-493-9473
15	662.5	33.4	421	15	US-10-369-493-9262

16	583	29.4	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	452.5	22.8	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	271.5	13.7	353	15	US-10-369-493-15733	Sequence 15733, A
19	271.5	13.7	353	15	US-10-369-493-16119	Sequence 16119, A
20	271.5	13.7	582	10	US-09-927-827-48	Sequence 48, Appl
21	247.5	12.5	533	12	US-10-424-599-176232	Sequence 176232, A
22	243	12.2	375	15	US-10-369-493-12537	Sequence 12537, A
23	235	11.8	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	196.5	9.9	438	12	US-10-425-114-66753	Sequence 66753, A
25	194.5	9.8	314	15	US-10-369-493-15726	Sequence 15726, A
26	194.5	9.8	314	15	US-10-369-493-16112	Sequence 16112, A
27	194.5	9.8	337	15	US-10-369-493-15359	Sequence 15359, A
28	188	9.5	517	12	US-10-107-431-259	Sequence 259, App
29	187	9.4	644	16	US-10-437-963-139883	Sequence 139883, A
30	182.5	9.2	917	16	US-10-437-963-181484	Sequence 181484, A
31	179.5	9.0	346	10	US-09-917-378-8	Sequence 8, Appli
32	175.5	8.8	375	10	US-09-917-378-3	Sequence 3, Appli
33	175.5	8.8	375	10	US-09-917-378-6	Sequence 6, Appli
34	175.5	8.8	356	10	US-09-917-378-1	Sequence 1, Appli
35	168	8.5	356	10	US-09-917-378-7	Sequence 7, Appli
36	164.5	8.3	558	16	US-10-437-963-185049	Sequence 185049, A
37	158.5	8.0	518	10	US-09-769-734-56	Sequence 56, Appl
38	154.5	7.8	722	12	US-10-282-122A-51993	Sequence 51993, A
39	153.5	7.7	1449	16	US-10-437-963-176100	Sequence 176100, A
40	152.5	7.7	533	12	US-10-424-599-150988	Sequence 150988, A
41	151.5	7.6	356	15	US-10-369-493-17756	Sequence 17756, A
42	149.5	7.5	268	12	US-10-424-599-284543	Sequence 284543, A
43	149.5	7.5	499	12	US-10-424-599-249378	Sequence 249378, A
44	148	7.5	217	16	US-10-437-963-112718	Sequence 112718, A
45	147.5	7.4	315	15	US-10-369-493-9177	Sequence 9177, Ap

ALIGNMENTS

RESULT 1

US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Magbool, Shahana B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE WHICH DEGRADE
; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981,900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match	99.6%	Score	1979	DB	9	Length	562
Best Local Similarity	99.7%	Pred. No.	5.4e-186				
Matches	357	Conservative	1	Mismatches	0	Gaps	0
Indels	0						
QY	1	AGGGWHTSGREILDANNVPVRIAGINWFGTCTNVVHGLWSRDYRMLQIKSLGNT	60				
Db	42	AGGGWHTSGREILDANNVPVRIAGINWFGTCTNVVHGLWSRDYRMLQIKSLGNT	101				
QY	61	IRLPYSDDLKPGTWPNSINFRQMDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC	120				
Db	102	IRLPYSDDLKPGTWPNSINFRQMDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC	161				
QY	121	SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWGGDPSIDWRL	180				
Db	162	SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWGGDPSIDWRL	221				

Db 140 YNSIPGGLVNGSNOAAIKEKYQVQOQIATKFSNVRERLIFESMNEVPD---GNYGNP 196
Qy 175 SIDWRLAERAGNAVLSVNPILL--IFVEGVQSY--NGDSYW-----WGNLQ-GAGQYP 224
Db 197 -----NRAAYANLNAYNQIFVDTVROTGGNNARWLLIPGWNTHIDYTVGNYG 244
Qy 225 VVL-----NVFN---RLVYSAHDY-----ATSVYP---QTFWSDP 253
Db 245 PALPTDHFPSAIPSSOKRINISAHYSPWDPAGEENGNIQWGAATNPSSKSTWQED 304
Qy 254 TFPNNMFGIWNK--NWGYLFNQNIAPVWLGEFGT---TLOSTTDQWKLKTLVQYLAPTA- 307
Db 305 YLNAQFKSMYDKFVTQGY-----PVGIFGSGIDKTAIDYSTNNVYRQAYAKAVTATAK 357
Qy 308 QYGADSFQWTFWSNPNDSGDTG 329
Db 358 KYGAVPVYVW-----DNHGNG 372

Search completed: August 2, 2004, 16:38:35
Job time : 15.6667 secs

Query Match 7.2%; Score 143; DB 4; Length 490;
Best Local Similarity 23.1%; Pred. No. 3.2e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;
US-09-339-159B-2

Qy 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHG-LWSRDY-RSMLDQIKSLGY 58
Db 32 ANSGF-YVSGTLLDANGNPFVNRGIN-----HGHAWKDQATTATEGIANTGA 79
Qy 59 NTRLPYS-----DDILKPGTMFNSINFRQNI-----QDLOGLTSLQVMDKIYAY 103
Db 80 NTVRIVLSGGQWTKDDI---HTVRNLISLAEDNHLVAVPEVHDATGYDSIASLRAVDY 136
Qy 104 AQQIGLRIILDRHRPDCSGQSALWYTSVSEATWISDLQALAQRYKGNPTVVVGFDLHNE- 162
Db 137 -----WIEMRSALI-----GREDTVIINIANEW 159
Qy 163 --PHDPACWCGDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYNGSYWGNLQGA 220
Db 160 FGSWEGDAWADG---YKQAIPLRNA--GLNHTLMV-----DAAGN----- 195
Qy 221 GOYEVVLNPNRLVYSAHDYATSVY---PQ--TWFSDDTTP---NNMFGIWNKNWGYLPN 272
Db 196 GQPP-----QSIHDIQREVFNADFORNTMFSIHMYEYAGGNASQV-RTNIDRVLN 244
Qy 273 QNIAPVWLGEFGTTLQSTTDQWKLKTLVQYLREPTAQYAGDSFQWTFWSNPNDSGDTGGIL 332
Db 245 QDLALV-IGFEGH--RHTNGVDVEATIMSY---SEQRGVG---WLAWSWKNGG----- 288
Qy 333 KDDWQVTVDTVKD 344
Db 289 -PEWEYLDLSND 299

RESULT 15
US-09-797-464A-4
; Sequence 4, Application US/09797464A
; Patent No. 6630340
; GENERAL INFORMATION:
; APPLICANT: Wilting, Reinhard
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Schilein, Martin
; TITLE OF INVENTION: Family 5 Xyloglucanases
; FILE REFERENCE: 6073.200-US
; CURRENT APPLICATION NUMBER: US/09/797,464A
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Paenibacillus sp.
US-09-797-464A-4

Query Match 7.2%; Score 142; DB 4; Length 400;
Best Local Similarity 22.5%; Pred. No. 3e-05;
Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;
US-09-339-159B-2

Qy 3 GGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLOIKSLGYNTIR 62
Db 48 GAGWNLGNQLEATVNGVPSETA---W-----GNPW-----TPELIKKVKAAAGFKTIR 92
Qy 63 LP--YSDDLKPGTMFN-SINFROMNODLOGLTSLQVMDKIYAYAGQIGLRIILDRHRPD 119
Db 93 IPVSYLNIH---GAPNYTTINAAWLN-----VQTVVDYAYNEGLVYVNIHGDG 139
Qy 120 CSQSALWYTSVSEATWISD-----LQALAQRYKGNPTVVVGFDLHNEPHDPACWCGD 174

```

US-08-525-697-2
; Sequence 2, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Helat-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-525-697-2
Query Match 8.9%; Score 176.5; DB 1; Length 377;
Best Local Similarity 23.1%; Pred. No. 1.3e-08;
Matches 81; Conservative 55; Mismatches 123; Indels 91; Gaps 19;

QY 24 AGIN--WFGFTCNVYVHGLWSRYRSMQDKSLGYNVTLRLPYSDDLKPGTTPNSINF 81
DB 50 AGTNSYMGFTLNDDVDLWMSQAASDLKILRVMGFNDVNTKPTD-----GTVVYQLHA 104
QY 82 RQMNQDLQGLTSQVMKIVAYAGQIGLRILD--RHRPDCSGQSAL-----WY 128
DB 105 NGTSTINTGADQLRLDYVTVSAEKYVKLIINFVNEWTYGGMQAVYVAYGAAQTDYF 164
QY 129 TSSVSEATWISDLQALQARYKGNVTYVGFDLHNEPHDPACWGCDPSIDWRLAAERAGNA 188
DB 165 TNTAIQARYKNIYKAVVSRYSSAAIFAWELANEPR---CQGC-DTSVLYNWISD-TSKY 219
QY 189 VLSVNEPLLIFVEGVQSVNGSYWVGNGNLQAGQYPVV----LNVPNRLVYSAHDYAT-S 243
DB 220 IKSLDSKHLVTI-----GDE-GFGLDVDSGYPPTYGEGNFTKNLGIITDGTGLH 271
QY 244 VYPQTFSDPTFPNMPGINKNKGWYLFNQINIA-----PVMIGFEGTT-----LQSTT 291
DB 272 LYPDWSGTSYD-----WGNW-----ITAHAAACKAVGPKCLLEBYGVTSNHCIVESPW 320
QY 292 DCT-----WLKTLVQYLRPTAQYGADSFQWTF--WSWNPDSGDT 328
DB 321 QQTAGNATGISGDLYW-----QYGT-----TFSWQGSFNDGNT 353

```

```

RESULT 13
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schmort, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4
Query Match 7.2%; Score 143; DB 4; Length 476;
Best Local Similarity 23.1%; Pred. No. 3.1e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDY-RSMLDQIKSLGY 58
DB 1 ANSGF-YVSGTTLTDANGNPFVWRCIN-----HGHAWYKQDQATTALEGIANTGA 48
QY 59 NTRLPYS-----DDILKPGTTPNSINFQMN-----ODLQGLTSLQVMDKIVAY 103
DB 49 NTVRIVLSDGGQWTKDDI---HTVERNLSLAEDNHLVAVPEVHDATGYDSTIASLNRAVDY 105
QY 104 AGQIGLRILDRHRPDCSGQSALWYTSVSEATWISDLQALQARYKGNPTVVGFPLHNE- 162
DB 106 -----WIEMRSALI-----GKEDIVINIANEW 128
QY 163 --PHDPACWGCDPSIDWRLAAERAGNAVLVSNPNLLIFVEGVQSVNGSYWVGNGNLQGA 220
DB 129 FGSWEGDAWAG-----YKQATPLRNA--GLNHTLMV-----DAAGW----- 164
QY 221 GQYPVVLNVPNRLVYSAHDYATSVY---PQ--TWESDPTFP---NNMFGINKNKGWYLFN 272
DB 165 GQFP-----QSIHDYGREVFNADPQNTMFSIHMEYAGGNASQV-RTNIDRVIN 213
QY 273 QNIAPVWLGEFGTTTLQSTTDTQTLKTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGIL 332
DB 214 QDLALV-IGEFCH--RHINGDVDEATIMSY---SEQRGVG---WLAWSWKNG----- 257
QY 333 KDDWQTVDTVKD 344
DB 258 -PEWEYLDLSD 268

RESULT 14
US-09-339-159B-2
; Sequence 2, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schmort, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2

```

```
QY 338 TVDTVKDGYLAP 349
Db 562 TWDEQKYNFLKP 573

RESULT 10
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ADDRESSEE: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43
Query Match 31.4%; Score 623; DB 3; Length 1426;
Best Local Similarity 35.5%; Pred. No. 1.6e-50;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNVPVRIAGINWFGFTCNVYVHGLMSRDYRSMDOIKSLGYNTIRLPYSDDI 69
Db 1022 GNKIVDKGKPVWLTGVNWFNGFTNTVDFGVWSCNLKSLAIAIARGFNLLRPISAEI 1081
QY 70 L---KPGTMPN--SINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRH--RPDCSQ 123
Db 1082 ILNWSKGIYKPNINY--YVNPLEGLTSLVDFVVKTKVEGLKIMLDIHSAKTDAMGH 1140
QY 124 -SALWYTSVSEATWISDLQALQARYKGNVTGVDFLHNEPHDP-----ACWGCGRS 175
Db 1141 IYPVWYTDITTPEDYKACEWITERYKNDTDITVAFDLKNEPHGKPMQDSVFAKWDNSTDI 1200

QY 176 IDWRLAARAGNAVLNVNPNLLIFVGVQSYNGD-----SYWVGGNLQAGQ 222
Db 1201 NNWKYAAETCAKRIKAPNPNLLIVIEIAYPKDDVTWTSSKSSDYSTWVGNLKGVK 1260
QY 223 YPVVL-NVPNRLVYSAHDYATSVYPTW-----ESDPTFPNNMFGIWNKNWGYLFNQNIAP 277
Db 1261 YPINLGOYQKVVYSPHDYGLVYQQPWFYPTKOTLYND---CWRDNWYIMDNGIAP 1317
QY 278 VWLGEFGTTIQTSDQTLWKLIVQYLRPTAQYQADSFQWTFWSNPNDSGDTGGILKDDWQ 337
Db 1318 LLIGEWGGYLDGGDNEKWMYLRDYII-----ENHIHTFWCYNANSNGDTGGLVGYDFS 1371
QY 338 TVDTVKDGYLAP 349
Db 1372 TWDEQKYNFLKP 1383

RESULT 11
US-08-276-213-2
; Sequence 2, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Laymon, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: N-terminal
US-08-276-213-2
Query Match 10.3%; Score 205; DB 1; Length 38;
Best Local Similarity 97.4%; Pred. No. 8.2e-13;
Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFTCNVYV 38
Db 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFTCNVYV 38

RESULT 12
```

Db 199 TDWNLAAER:GKAILKVAPHWLFVEGTQTNPKTSSYKGNANWGNLMAVKDYPV- 257
Qy 227 LNVP-NRLVYSADHYATSVYPTQWFSDDT-FPNMFGIWNKNGYLFNQNIAPVWLGEGF 284
Db 258 -NLPRNKLAVSPHYGPDVYNQPYFGPAKGFPPNLPDIWYHFGYVYKLELGYGAVWIGEGF 316
Qy 285 TTL---QSTTDQWTKLVOYLRTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWTV 339
Db 317 GKYGHGDPDRVIMQNKLVDMW--IENKFCDFP---YWSWNPDSGDTGGILQDDWTI 369

RESULT 8

US-09-869-197-5
; Sequence 5, Application US/09869197
; Patent No. 6566113
; GENERAL INFORMATION:
; APPLICANT: TAKAYAMA, Masanori
; APPLICANT: UMEDA, Kahoko
; APPLICANT: KOYAMA, No. 6566113uto
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: TAKAYAMA-6
; CURRENT APPLICATION NUMBER: US/09/869,197
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/JP99/07009
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: JP 366237/1998
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii OT3
US-09-869-197-5

Query Match 38.4%; Score 762; DB 4; Length 458;
Best Local Similarity 45.0%; Pred. No. 1.4e-64;
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSREILDANNVPRIAGINWFGPCTCNVYVHGLWSRDSYRMLDQIKSLGYNTIRLPYSD 67
Db 54 TSREE-----TPHFGVNWFGPCTCNVYVHGLWKRKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFRQMNQDQGLTSQVMDKIVAVAGQIGLRILDRHRPDCGQSALW 127
Db 108 ESKVPGTQPIDYKX-NPDLRGDLSLIQIMEKIKKAGDIGFVLLDYHRIQCTHIEPLW 166
Qy 128 YTSVSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEFSEEDFTINTWIEVAKRFGKYNWVIGADLKNEPHSVTSPPAAVYTDGTGATWGMGNEA 226
Qy 176 IDWRLAARAGNAVLSPNPLLIIFVEGVQSYN-----GDSYMWGNLQAGQYPPV 226
Db 227 TDWNLAAER:GKAILKVAPHWLFVEGTQTNPKTSSYKGNANWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSADHYATSVYPTQWFSDDT-FPNMFGIWNKNGYLFNQNIAPVWLGEGF 284
Db 286 -NLPRNKLAVSPHYGPDVYNQPYFGPAKGFPPNLPDIWYHFGYVYKLELGYGAVWIGEGF 344
Qy 285 TTL---QSTTDQWTKLVOYLRTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWTV 339
Db 345 GKYGHGDPDRVIMQNKLVDMW--IENKFCDFP---YWSWNPDSGDTGGILQDDWTI 397

RESULT 9

US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy W.
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 31.4%; Score 623; DB 3; Length 616;
Best Local Similarity 35.5%; Pred. No. 4.6e-51;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;
Qy 10 GREILDANNVPRIAGINWFGPCTCNVYVHGLWSRDSYRMLDQIKSLGYNTIRLPYSDDI 69
Db 212 GNKIVDKGPKVWLGVNWFNTGTVDFGVSNLKSALAEIANRGNLRLVPISAEL 271
Qy 70 L---KPGTMPN-SINFRQMNQDQGLTSQVMDKIVAVAGQIGLRILDRH--RPDCSGQ 123
Db 272 ILNWSKGIYKPNINY-VVNPELGTSLEVFDFVVKTCKEVGLKIMLDIHSAKTDAMGH 330
Qy 124 -SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----ACWCGDPS 175
Db 331 YPVWYTTIIPEDYKACEWITERYKNDTIVAFDLKNEPHGPKWPQSVFAKNDNSTDI 390
Qy 176 IDWRLAARAGNAVLSPNPLLIIFVEGVQSYNGD-----SYMWGNLQAGQ 222
Db 391 NNWYAAETCAKRLAKPNMLIVIEGIEAYPKDDVTWTSKSSDYSTWNGNLRGVK 450
Qy 223 YPVVL-NVFNRLVYSADHYATSVYPTQW----FSDPTFPNMPGWNKNGYLFNQNIAP 277
Db 451 YPINLGQYQNKVYSPHYGPLVYQQPWFYQFPFTKDTLYND---CWRDNWTYIWDNGIAP 507
Qy 278 VWLGEFTTQSTTDQWTKLVOYLRTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQ 337
Db 508 LLIGWGGYLDGDNEXKWTYLRDVI-----ENHIHTFWCYNANSGDTGGLVGYDFS 561

Query Match 38.8%; Score 770; DB 3; Length 553;
Best Local Similarity 45.2%; Pred. No. 3.3e-65;
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPVRIAGINWFGFETCYVYVHGLWSRDRYRMLDQIKSLGYNTRLRPLYSDDLKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPFCQSVKP 63

QY 73 GTWPNINFRQNMQDLOGLTSLQVMDKIYAYAGQIGLRIILDRHRPDCSGSALWYTSV 132
DB 64 GTMPTAIDYAK-NPDQLGSDSVQIMEKLIKAGDLGIFVLLDYHRIGCNFIEPLWYDTSF 122

QY 133 SEATWISDLQALQARYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180
DB 123 SEQDYINTWVEAQRFGKYWNVIGADLKNEPHSSPAPAAAYTDCSGATWGNNAIDWNL 182

QY 181 AERAGNAVLNVPNLLIFVEGVQ-----SYN-GDSYVWGMQLQAGQYVVLNVP- 230
DB 183 AAEIRGRAILEVAPQWVIFVEGTQFTTPBIDGRYKGNHAWGNLGMVRYKYPV--NLPR 240

QY 231 NRVLSAHDYATSVYPTWFSDF--TFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTL- 287
DB 241 DKVYSPQVYGVSEVYDQYF-DPCEGFPDNLPEIWHHFGYVKLDLGYFVVIIEFGGKYG 299

QY 288 --QSTTDTWLTVLQVLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSWNPNSGDTGGILKDDWTTI 348

RESULT 6
US-09-430-669-2
; Sequence 2, Application US/09430669
; Patent No. 6329187
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,669
; FILING DATE: 28-Oct-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: 22-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: <Unknown>
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-430-669-2

Query Match 38.8%; Score 770; DB 4; Length 553;
Best Local Similarity 45.2%; Pred. No. 3.3e-65;
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPVRIAGINWFGFETCYVYVHGLWSRDRYRMLDQIKSLGYNTRLRPLYSDDLKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPFCQSVKP 63

QY 73 GTWPNINFRQNMQDLOGLTSLQVMDKIYAYAGQIGLRIILDRHRPDCSGSALWYTSV 132
DB 64 GTMPTAIDYAK-NPDQLGSDSVQIMEKLIKAGDLGIFVLLDYHRIGCNFIEPLWYDTSF 122

QY 133 SEATWISDLQALQARYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180
DB 123 SEQDYINTWVEAQRFGKYWNVIGADLKNEPHSSPAPAAAYTDCSGATWGNNAIDWNL 182

QY 181 AERAGNAVLNVPNLLIFVEGVQ-----SYN-GDSYVWGMQLQAGQYVVLNVP- 230
DB 183 AAEIRGRAILEVAPQWVIFVEGTQFTTPBIDGRYKGNHAWGNLGMVRYKYPV--NLPR 240

QY 231 NRVLSAHDYATSVYPTWFSDF--TFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTL- 287
DB 241 DKVYSPQVYGVSEVYDQYF-DPCEGFPDNLPEIWHHFGYVKLDLGYFVVIIEFGGKYG 299

QY 288 --QSTTDTWLTVLQVLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSWNPNSGDTGGILKDDWTTI 348

RESULT 7
US-09-869-197-1
; Sequence 1, Application US/09869197
; Patent No. 6566113
; GENERAL INFORMATION:
; APPLICANT: TAKAYAMA, Masanori
; APPLICANT: UMEIDA, Kahoko
; APPLICANT: KOYAMA, No. 6566113uto
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoskin
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: TAKAYAMA-6
; CURRENT APPLICATION NUMBER: US/09/869,197
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/JP99/07009
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: JP 366237/1998
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 1
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii OT3
US-09-869-197-1

Query Match 38.4%; Score 762; DB 4; Length 430;
Best Local Similarity 45.0%; Pred. No. 1.3e-64;
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFETCYVYVHGLWSRDRYRMLDQIKSLGYNTRLRPLYS 67
DB 26 TSGEE-----TPHLFGVNWFGFETPNVYVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 79

QY 68 DILKPGTNPNSINFRQNMQDLOGLTSLQVMDKIYAYAGQIGLRIILDRHRPDCSGSALW 127
DB 80 ESKVEGTQPIGIDYSK-NPDLRGLDSLQIMEKLIKAGDLGIFVLLDYHRIGCTHIEPLW 138

QY 128 YTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
DB 139 YTEDESEEDFINTEVAKRFGKYWNVIGADLKNEPHSVTSPAAAYTDCGTATWGMGNA 198

QY 176 IDWRLAERAGNAVLNVPNLLIFVEGVQSYN-----GDSYVWGMQLQAGQYV 226

Query Match 38.8%; Score 770; DB 1; Length 553;
Best Local Similarity 45.2%; Pred. No. 3.3e-65;
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPIRIAGNWFETCNVYVHGLWSRDYRSLDQIKSLGYNTRILRPSDDILKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGYNTRILRPSCTQSKVP 63
QY 73 GTPMNSINFRQMNQDLQGLTSLOWMDKIVAYAGQIGLRIILDRHRPDCSQSALWTSSV 132
DB 64 GTMPTAIDYAK-NPDQLGLDSVQIMEKIIKKAGDLGIFVLDDYHRIKCNFIEPLWTDGSF 122
QY 133 SEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCDDPSIDWRL 180
DB 123 SEQDYINTWVEVAQRFGKYNNVIGADLNKNEPHSSSPAPAAAYTDCSGATWGMGNATDNWL 182
QY 181 AAERAGNAVLNVNPNLLIFVEGVQ-----SYN-GDSYWMGNLQAGQYPPVNLNVP- 230
DB 183 AAERIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWNHNAWGNLGMVKYKYPV--NLPR 240
QY 231 NRVYSADHYATSYPOTWFSDDP--TPNNMFGIWNKNWGYLFNQNIAPVWLGEFTTL- 287
DB 241 DKVYSPQVYGVSEYVDQYF-DPGEFGPDNLPEIWHHFGYVLDLGYPPVIVGEGKYG 299
QY 288 --QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDFP---YWSWNPNSGDTGGILKDDWTTI 348

RESULT 4

US-09-066-544-2
; Sequence 2, Application US/09066544
; Patent No. 6001984
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,544
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28, 019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-066-544-2

Query Match 38.8%; Score 770; DB 3; Length 553;
Best Local Similarity 45.2%; Pred. No. 3.3e-65;
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPIRIAGNWFETCNVYVHGLWSRDYRSLDQIKSLGYNTRILRPSDDILKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGYNTRILRPSCTQSKVP 63
QY 73 GTPMNSINFRQMNQDLQGLTSLOWMDKIVAYAGQIGLRIILDRHRPDCSQSALWTSSV 132
DB 64 GTMPTAIDYAK-NPDQLGLDSVQIMEKIIKKAGDLGIFVLDDYHRIKCNFIEPLWTDGSF 122
QY 133 SEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCDDPSIDWRL 180
DB 123 SEQDYINTWVEVAQRFGKYNNVIGADLNKNEPHSSSPAPAAAYTDCSGATWGMGNATDNWL 182
QY 181 AAERAGNAVLNVNPNLLIFVEGVQ-----SYN-GDSYWMGNLQAGQYPPVNLNVP- 230
DB 183 AAERIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWNHNAWGNLGMVKYKYPV--NLPR 240
QY 231 NRVYSADHYATSYPOTWFSDDP--TPNNMFGIWNKNWGYLFNQNIAPVWLGEFTTL- 287
DB 241 DKVYSPQVYGVSEYVDQYF-DPGEFGPDNLPEIWHHFGYVLDLGYPPVIVGEGKYG 299
QY 288 --QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDFP---YWSWNPNSGDTGGILKDDWTTI 348

RESULT 5

US-08-951-086-2
; Sequence 2, Application US/08951086
; Patent No. 6074867
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,086
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/651,572
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28, 019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-951-086-2

Query Match 99.6%; Score 1979; DB 1; Length 358;
Best Local Similarity 99.7%; Pred. No. 2.5e-181;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGIIGLRIILDRHRPDC 120
DB 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGIIGLRIILDRHRPDC 120

QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVNLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVNLNPNRLVYSAHDY 240

QY 241 ATSVYPTWFSPTFFNNMFGIWNKNGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300
DB 241 ATSVYPTWFSPTFFNNMFGIWNKNGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300

QY 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2
US-08-276-213-3
; Sequence 3, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Layton, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 99.6%; Score 1979; DB 1; Length 521;
Best Local Similarity 99.7%; Pred. No. 4.4e-181;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGIIGLRIILDRHRPDC 120
DB 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGIIGLRIILDRHRPDC 120

QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVNLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVNLNPNRLVYSAHDY 240

QY 241 ATSVYPTWFSPTFFNNMFGIWNKNGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300
DB 241 ATSVYPTWFSPTFFNNMFGIWNKNGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300

QY 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3
US-08-651-572-2
; Sequence 2, Application US/08651572
; Patent No. 5789228
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herion
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-651-572-2

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds
(without alignments)
1260.143 Million cell updates/sec

Title: US-09-997-504A-14
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDFV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	358	1	US-08-604-913B-11
2	1979	99.6	521	1	Sequence 11, Appli
3	770	38.8	553	1	Sequence 3, Appli
4	770	38.8	553	1	Sequence 2, Appli
5	770	38.8	553	3	Sequence 2, Appli
6	770	38.8	553	3	Sequence 2, Appli
7	762	38.4	430	4	Sequence 2, Appli
8	762	38.4	430	4	Sequence 1, Appli
9	623	31.4	616	3	Sequence 5, Appli
10	623	31.4	1426	3	Sequence 47, Appli
11	205	10.3	38	1	Sequence 43, Appli
12	176.5	8.9	377	1	Sequence 2, Appli
13	143	7.2	476	4	Sequence 2, Appli
14	143	7.2	490	4	Sequence 4, Appli
15	142	7.2	400	4	Sequence 2, Appli
16	138	6.9	348	4	Sequence 7, Appli
17	138	6.9	468	4	Sequence 4, Appli
18	138	6.9	468	4	Sequence 4, Appli
19	138	6.9	468	4	Sequence 4, Appli
20	138	6.9	468	4	Sequence 8, Appli
21	138	6.9	493	4	Sequence 2, Appli
22	138	6.9	493	4	Sequence 2, Appli
23	138	6.9	493	4	Sequence 2, Appli
24	138	6.9	493	4	Sequence 6, Appli
25	137	6.9	395	4	Sequence 2, Appli
26	135	6.8	363	4	Sequence 11, Appli
27	132.5	6.7	663	4	Sequence 61, Appli

28	132.5	6.7	680	4	US-09-134-078-25
29	131	6.6	24	1	US-08-276-213-1
30	129	6.5	331	4	US-09-339-159B-12
31	126	6.3	320	4	US-09-339-159B-22
32	120	6.0	331	3	US-08-849-751-4
33	120	6.0	331	4	US-08-478-816-4
34	114.5	5.8	429	1	US-08-745-977-4
35	114.5	5.8	429	3	US-09-040-699A-4
36	112.5	5.7	360	4	US-09-134-078-27
37	110.5	5.6	461	1	US-08-672-571A-3
38	110.5	5.6	490	1	US-08-672-571A-1
39	109	5.5	317	2	US-09-066-075-2
40	109	5.5	317	2	US-08-518-615A-2
41	109	5.5	317	3	US-08-951-889-2
42	109	5.5	317	3	US-09-472-857-2
43	108	5.4	327	2	US-08-169-948B-16
44	108	5.4	327	2	US-08-448-873-16
45	108	5.4	327	3	US-08-382-452D-16

ALIGNMENTS

RESULT 1
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.
; APPLICANT: Himmel, Michael E.
; APPLICANT: Baker, John O.
; APPLICANT: Chou, Yat-Chen
; TITLE OF INVENTION: METHOD FOR INCREASING
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASC II (DOS) text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/604,913B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,213
; FILING DATE: 15-070-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Edna M. O'Connor
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: 95-56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/384-7573
; TELEFAX: 303/384-7499
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; FEATURE:
; NAME/KEY: E1-CAT

US-08-604-913B-11

the invention is a thermostable enzyme which is heat stable, catalyses the enzymatic hydrolysis of cellulose and is able to renature and regain activity after exposure to temperatures of 60-105 degrees Centigrade. The polynucleotide and polypeptide can be used to raise an antibody, generate a variant and as a probe for isolating or identifying other endoglucanase genes. Also disclosed is a method for comparing a first sequence to a reference sequence, and for identifying a feature, through the use of a computer program. The endoglucanase has increased activity and stability at increased pH and temperature. The sequence presented is the *T. maritima* endoglucanase protein

Sequence 553 AA;

	Query Match	38.8%; Score 770; DB 5; Length 553;
	. Best Local Similarity	45.2%; Pred. No. 9.le-64;
	Matches	160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;
QY	13	ILDANNVPVRIAGINWGFETCNVVHGLSRDYSRLDOIKS LGYN TIRLPYP SDDILKP 72
DB	4	VATGEETPIHLFGVNWGFETPNVVHGLSRNWE DMLLQIKSLGFINAIRLPFTCQVKP 63
QY	73	GTPENSINFQMNDLOQLTSLQVM DKIVAVAGOIGLRILDRHRPCSGOSALWYTSSV 132
DB	64	GTWPTAIDYAK-NFDLQGLDSVQIMEXIIKKAGDIGIFVLVDHYRIGNFIERPLWYTDSP 122
QY	133	SEATWISDLQALAQR YKGNPVWGFDLHNEBHP-----ACWGGCDSPSIDWL 180
DB	123	SEQDYINTWVEVAQRF GKYNMVIADLKNEPHSSSPAAYTDGSGATGMGNNA TDNL 182
QY	181	AAERAGNAVLSVNELLIFEGVQ-----SYN-GDSYWMGNLQAGAGVPV LNV P - 230
DB	183	AAERIGRAILEVAPOWIFVEGTOTTFEIDGRYKMGHNWNGNLGMGRKY PV - NLPR 240
QY	231	NRLVYSRAHDVATS VYPOTWFSDP--TPENNPMGIWNKNWGYLFNQNTAPVWLGBFGTTL- 287
DB	241	DKVYVSQVYGSEVYDQYPF-DPEGEFPDNLPELWYHHFGYGVKL DLGYPPVVGEGGKYG 299
QY	288	--QSTDTQTLTKLVQLRPTAQYCAQS FQTWFSWNPDSGD TGILKDMMQT 339
DB	300	HGGPRPVTWKNI IDW--IONKFCDFP---YWSWNPENSGDTGILKDMMTTI 348

RESULT 12	
AAAB10344	
ID	AAAB10344 standard; protein; 430 AA.
XX	
XX	
XX	AAAB10344;
XX	
DT	24-NOV-2000. (first entry)
XX	
XX	P. horikoshii OT3 cellobiohydrolase associated protein.
DE	
XX	
XX	Cellobiohydrolase; poly(D-glucopyranose) decomposition; glucose;
KW	cellulose breakdown.
KW	
XX	
OS	Pyrococcus horikoshii.
XX	
XX	WC2000039288-A1.
FN	
XX	
XX	06-JUL-2000.
PD	
XX	
XX	
PF	14-DEC-1999; 99WO-JF007009.
XX	
PR	24-DEC-1998; 98JP-00366237.
XX	
XX	(TAKI) TAKARA SHUZO CO LTD.
PA	
XX	
PI	Takayama M, Umeda K, Koyama N, Asada K, Kato I;
XX	
DR	WPI; 2000-452391/39.
DR	N-PSDE; AAA71327.
XX	
XX	
PT	Polypeptides with heat-resistant cellobiohydrolase activity for efficient

PT breakdown of cellulose biomass.
XX
XX
PS Claim 1; Page 41-43; 50pp; Japanese.
XX
CC This invention describes a novel polypeptide originating in *Pyrococcus*
CC *horikoshii* OT3 which has cellobiohydrolase activity. The polypeptide of
CC the invention is capable of decomposing poly(D-glucopyranose) having beta
CC -1,4 bonds and can be used for the efficient and straightforward
CC breakdown of cellulose biomass to glucose. This sequence represents the
CC P. horikoshii OP3 cellobiohydrolase associated protein described in the
CC method of the invention
XX
XX Sequence 430 AA;
SQ

Query Match	38.4%;	Score 762;	DB 3;	Length 430;
Best Local Similarity	45.0%;	Pred. No. 3.6e-63;		
Matches 161;	Conservative 48;	Mismatches 109;	Indels 40;	Gaps 10;
Qy	8	TSGREILDANNVPRIAGINWGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLPYS	67	
Db	26	TSGE-----TPIHLFGVNWGFETPNVHVHGWKRWEDMLLQIKSLGFNAIRLEFCT	79	
Qy	68	DILKPGTMPNSINFRQWQDLOGJTSLQVMDKIVAYAGQIGLRIILDRHRPCDSCQSALW	127	
Db	80	ESVKPGTQPIGIDYSK-NPDLRGLDSLSIQWEKIIKKAAGDLGIFVLLDYHRIGCTHIEPLW	138	
Qy	128	YTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGCDPS	175	
Db	139	YTFDFSEDFINTWIEVAKRFGKYWNVIGADLKNEPHSVTSPAAAYTDGTGATWGMGNPA	198	
Qy	176	IDWELAAERAGNAVLSNPNNLIIFVEGVQSYN-----GDSYMWGNLQAGAQYPVV	226	
Db	199	TDWNLAAERIGKAILKVAPHLIFVEGTQTNPKTSSYKNGYNAMWGNLMAVKYPV-257		
Qy	227	LNVP-NELVYSADHYATSVYPQWFSDPT-FPNNMPGIMWNKNXGYLEFNQNIAPVWLGEFG	284	
Db	258	NLPRNKLVSYPHYGPDVYNQYVFGPAKGFPDNLPDIIWVHFRFGYVKLELGYSVWIGFG	316	
Qy	285	TTL---QSTTDQWTKTLVLOYLRPTAQYAGDASQWTFWSWNPDSGDTGGTILKDDWQIV	339	
Db	317	GKYGHGGDPRDVTWONKLVDMW---IENKFCDF---YWSWNPDSGDTGGHILQDDWTTI	369	

RESULT 13	
AAAB10345	
ID	AAAB10345 standard; protein; 458 AA.
XX	
AC	AAAB10345;
XX	
DT	24-NOV-2000 (first entry)
XX	
DE	P. horikoshii O73 cellobiohydrolase associated protein #2.
XX	
KW	Cellobiohydrolase; poly(D-glucopyranose) decomposition; glucose;
KW	cellulose breakdown.
XX	
OS	Pyrococcus horikoshii.
XX	
FN	WO2000039288-A1.
XX	
PD	06-JUL-2000.
XX	
PF	14-DEC-1999; 99WO-JP007009.
XX	
PR	24-DEC-1998; 98JP-0036237.
XX	
PA	(TAKI) TAKARA SHUZO CO LTD.
XX	
PI	Takayama M, Umeda K, Koyama N, Asada K, Kato I;
XX	
DR	WPI; 2000-452391/39.
DR	N-PSDB; AAA71330.
XX	

XX Takayama M, Umeda K, Koyama N, Asada K, Kato I;
PI
XX WPI; 2000-452391/39.
DR N-PSDB: AAA71330.
DR

```

Db 183 AAEIRGRAILEVAPQWVIFVEGTQFTTPEIDGRYKKGHNWAGNLMGVKRYFV--NLPR 240
Qy 231 NRLVYSAHDYATSYPTWFSDDP--TFPNNMFGIKNNQWGYLFNQNIAPVWLGFEFTTL- 287
Db 241 DKLVSQVYGPVDYDQFYF-DPGEFDPNLPBIVYHHFGYVXLDLGPVWIGFGRKYG 299
Qy 288 --QSTTDQTLVQLRPTAQYGADSFQWTFWNNPDSGDTGGILKDDQTV 339
Db 300 HGGPRDVTWQNKIIDMW--IQNKFCDF--YMSWNPNSGDTGGILKDDWTII 348

RESULT 10
AAW34985
ID AAW34985 standard; protein; 841 AA.
XX AC AAW34985;
XX DT 21-AUG-2003 (revised)
XX DT 21-MAY-1998 (first entry)
XX DE Archaeobacterial thermostable endoglucanase.
XX KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
XX KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
XX KW archaeal bacterium; thermostable enzyme; thermophilic.
XX OS archaeon.
XX OS Unidentified.
XX PN WO9744361-A1.
XX PD 27-NOV-1997.
XX PF 22-MAY-1997; 97WO-US008793.
XX PR 22-MAY-1996; 96US-00651572.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PI Lam DE, Mathur EJ;
XX DR WPI; 1998-018435/02.
XX DR N-PSDB; AAT94193.
XX PT Endo:glucanase(s), preferably from archaeal bacterium, AEP11 la - useful
XX PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
XX PT bonds in cellulose.
XX PS Claim 1; Fig 1A; 164pp; English.
XX CC This protein comprises a thermostable 60.9 kDa endoglucanase of archaeal
XX CC bacterial strain AEP11a, a thermophilic isolate of a marine hydrothermal
XX CC vent. The endoglucanase is capable of degrading carboxymethylcellulose
XX CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It can be
XX CC produced from native cells or from recombinant host cells, especially
XX CC prokaryotic host cells transformed with a plasmid or virus-derived vector
XX CC including the endoglucanase DNA (see AAT94193). 23 Other, homologous
XX CC endoglucanases (see AAW34986-W35008) are also claimed. The endoglucanases
XX CC can be used to degrade cellulose for the conversion of plant biomass into
XX CC fuels and chemicals, for use in detergents, textiles, animal feed, waste
XX CC treatment, and in the fruit juice and brewing industries for the
XX CC clarification and extraction of juices. (Updated on 27-AUG-2003 to
XX CC correct OS field.)
XX SQ Sequence 841 AA;
XX Query Match 38.8%; Score 771; DB 2; Length 841;
XX Best Local Similarity 45.2%; Pred. No. 1.3e-63;
XX Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;
Qy 13 ILDANNVPRVRIAGNINWGFETCTVNVHGLWSRSDYRSMLDQIKSLGYNTILPYSDDLKP 72

```

```

Db 4 VATGETPIHLFGVNVFGFETPNVYVHGLWSRWNEDMLQIKSLGFNAIRLPFCTQSVKP 63
Qy 73 GTMENSINFRQWNOLOGLTSLQWMDXIVAYAGOIGLRILILDRHRPCSGQSALWYTSV 132
Db 64 GTMPTAIDYAK-NPDLOGLSDVQMEKLIKAGDLGIFVLIDYHRIKCNFTPELWYTDSP 122
Qy 133 SEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCSPSIDWRL 180
Db 123 SEQDINTWVEVAQRFEGKYVNVIGADLNKNEPHSSSPAPAAVTDCSGATWGNMNAIDWNL 182
Qy 181 AAEIRGRAILEVAPQWVIFVEGTQFTTPEIDGRYKKGHNWAGNLMGVKRYFV--NLPR 240
Db 183 AAEIRGRAILEVAPQWVIFVEGTQFTTPEIDGRYKKGHNWAGNLMGVKRYFV--NLPR 240
Qy 231 NRLVYSAHDYATSYPTWFSDDP--TFPNNMFGIKNNQWGYLFNQNIAPVWLGFEFTTL- 287
Db 241 DKLVSQVYGPVDYDQFYF-DPGEFDPNLPBIVYHHFGYVXLDLGPVWIGFGRKYG 299
Qy 288 --QSTTDQTLVQLRPTAQYGADSFQWTFWNNPDSGDTGGILKDDQTV 339
Db 300 HGGPRDVTWQNKIIDMW--IQNKFCDF--YMSWNPNSGDTGGILKDDWTII 348

RESULT 11
ABG70759
ID ABG70759 standard; protein; 553 AA.
XX AC ABG70759;
XX DT 09-DEC-2002 (first entry)
XX DE T. maritima endoglucanase protein.
XX KW Endoglucanase; enzyme; cellulose; plant cell wall; homopolysaccharide;
XX KW D-glucose; beta configuration; beta 1->4 glycosidic bond; hydrolysis;
XX KW 1,4-beta-D-glucan glucanohydrolase; exoglucanase;
XX KW 1,4-beta-D-glucan cellobiohydrolase; thermostable; catalysis; antigen.
XX OS Thermotoga maritima.
XX PN US2002120118-A1.
XX PD 29-AUG-2002.
XX PF 22-JUN-2001; 2001US-00888224.
XX PR 22-MAY-1996; 96US-00651572.
XX PR 24-APR-1998; 98US-00066544.
XX PR 28-OCT-1999; 99US-00430669.
XX PA (SHORT) SHORT J M.
XX PA (LAMD/) LAM D E.
XX PA (MATH/) MATHUR E J.
XX PI Short JM, Lam DE, Mathur EJ;
XX DR WPI; 2002-691213/74.
XX DR N-PSDB; AB554439.
XX PT New isolated nucleic acid encoding a polypeptide having endoglucanase
XX PT activity, useful as a probe for isolating or identifying other
XX PT endoglucanase genes having a sequence that is similar to its sequence.
XX PS Claim 35; Fig 5; 36pp; English.
XX CC The invention discloses an isolated nucleic acid encoding a polypeptide
XX CC from Thermotoga maritima which has an endoglucanase activity. Cellulose
XX CC is a tough, fibrous, water insoluble substance found in the cell walls of
XX CC plants. It consists of an unbranched homopolysaccharide of 10,000 to
XX CC 15,000 D-glucose units in a beta configuration linked by beta 1->4
XX CC glycosidic bonds. The enzymatic hydrolysis of cellulose requires the
XX CC action of both endoglucanases (1,4-beta-D-glucan glucanohydrolase) and
XX CC exoglucanases (1,4-beta-D-glucan cellobiohydrolase). The endoglucanase of

```


CC activity comprises replacing an active site-associated amino glycosyl-
CC stabilising amino acid with an amino acid that does not strongly bind a
CC disaccharide product in the active site. Conversely, the method for
CC making a glycosyl hydrolase with increased soluble substrate catalytic
CC activity comprises replacing a hydrophobic substrate-binding amino acid
CC with a positively charged residue. The invention also discloses mutants
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
CC produced according to the method of the invention. The Y245G mutant
CC (AAB48788) has improved activity with insoluble substrates, and the W42R
CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant
XX
SQ Sequence 521 AA;
Query Match 99.1%; Score 1969; DB 4; Length 521;
Best Local Similarity 99.4%; Pred. No. 2.5e-177;
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
QY 61 IRLPYSDDLKPGTNPNSINFRQNMNQLGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTNPNSINFRQNMNQLGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPVVLNPNRLVYSAHDY 240
DB 181 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTQWTFSDPTFPNNMPCGNKNGYLFNQNIAPVWLGEFGTTLQSTTDQTLWKLTV 300
DB 241 ATSVYPTQWTFSDPTFPNNMPCGNKNGYLFNQNIAPVWLGEFGTTLQSTTDQTLWKLTV 300
QY 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 7
AAB48786
ID AAB48786 standard; protein; 521 AA.
XX
AC AAB48786;
XX
DT 09-MAR-2001 (first entry)
XX
DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.
XX
KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
KW cellulose hydrolysis; ethanol production; fermentation; mutant; muten.
OS
OS Acidothermus cellulolyticus.
OS Synthetic.
XX
FN WO200070031-A1.
XX
PD 23-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-US013971.
XX
PR 19-MAY-1999; 99US-0134925P.
XX
PA (MIDE) MIDWEST RES INST.
XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TS, Thomas SR, Sakon J;
PI Decker SR;
XX WPI; 2001-061226/07.
DR
XX Preparation of glycosyl hydrolase with an increased catalytic activity on
PT insoluble substrate.
PT
XX Claim 16; Page 22-24; 30pp; English.
XX
XX The invention relates to a method for making glycosyl hydrolase mutants
CC with increased catalytic activity with either insoluble or soluble
CC cellulose substrates relative to the wild-type enzyme. The method for
CC making a glycosyl hydrolase with increased insoluble substrate catalytic
CC activity comprises replacing an active site-associated amino glycosyl-
CC stabilising amino acid with an amino acid that does not strongly bind a
CC disaccharide product in the active site. Conversely, the method for
CC making a glycosyl hydrolase with increased soluble substrate catalytic
CC activity comprises replacing a hydrophobic substrate-binding amino acid
CC with a positively charged residue. The invention also discloses mutants
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
CC produced according to the method of the invention. The Y245G mutant
CC (AAB48788) has improved activity with insoluble substrates, and the W42R
CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant
XX
SQ Sequence 521 AA;
Query Match 98.9%; Score 1965; DB 4; Length 521;
Best Local Similarity 99.4%; Pred. No. 6e-177;
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
QY 61 IRLPYSDDLKPGTNPNSINFRQNMNQLGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTNPNSINFRQNMNQLGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPVVLNPNRLVYSAHDY 240
DB 181 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTQWTFSDPTFPNNMPCGNKNGYLFNQNIAPVWLGEFGTTLQSTTDQTLWKLTV 300
DB 241 ATSVYPTQWTFSDPTFPNNMPCGNKNGYLFNQNIAPVWLGEFGTTLQSTTDQTLWKLTV 300
QY 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 8
ADD24922
ID ADD24922 standard; protein; 535 AA.
XX
AC ADD24922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Xanthomonas campestris cellulase #3.
XX
KW Directed genetic engineering; galactomannanase; reduced activity;
KW enhanced activity; xanthan gum production; suspension stability;
KW

DB 282 ATSVPTQWFSDFPTFPNNMGVWYLFNQNTAPVWLGEFGTTLQSTTDQWLKTLV 341
 QY 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTVTKDGYLAPIKSSIFDPV 358
 DB 342 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTVTKDGYLAPIKSSIFDPV 399

RESULT 5
 AAY69508
 ID AAY69508 standard; protein; 562 AA.
 XX
 AC AAY69508;
 XX
 DT 10-APR-2000 (first entry)
 DE
 DE Acidothermus cellulolyticus E1 endoglucanase.
 XX
 KW E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;
 KW beta-1,4-endoglucanase; endocellulase; thermostable.
 XX
 OS Acidothermus cellulolyticus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..41
 FT /note= "Putative signal peptide"
 FT Peptide 14..41
 FT /note= "Putative signal peptide (alternative)";
 FT Domain 42..404
 FT /note= "Catalytic domain"
 FT Region 405..460
 FT /note= "Linker region"
 FT Domain 461..562
 FT /note= "Cellulose binding domain (CBD)";
 XX
 CA2226898-A1.
 XX
 PD 25-SEP-1999.
 XX
 XX 25-MAR-1998; 98CA-02226898.
 XX
 PR 25-MAR-1998; 98CA-02226898.
 XX
 PA (MIDE) MIDWEST RES INST.
 XX
 PI Laymon RA, Adney WS, Thomas SR, Himmel ME;
 XX
 WPI; 2000-057663/08.
 DR N-PSDB; AAZ55924.
 XX
 PT Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
 PT for labeling or modifying a cellulose and for purifying or immobilizing a
 PT binding domain fusion protein to cellulose.
 XX
 PS Claim 1; Fig 2; 85pp; English.
 XX
 CC This sequence represents the Acidothermus cellulolyticus E1
 CC endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The
 CC cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which
 CC encode it are specifically claimed. The CBD is believed to be roughly
 CC wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge
 CC tip is inserted between the microfibrils of the cellulose fibre,
 CC disrupting the crystalline structure, and making the cellulose linkages
 CC more accessible to the catalytic domain of the E1 endoglucanase. The E1
 CC endoglucanase CBD is useful in labelling or modifying the surface of
 CC cellulose or other polysaccharides. Such modified cellulose can then be
 CC used in textile, pulp, paper, chemical and pharmaceutical industries.
 CC CBDs can be used in affinity purification of CBD-fusion proteins, and can
 CC also be used to immobilise the CBD-fusion proteins to a cellulose
 CC support. CBD-fusion proteins can be used to modify the chemical or
 CC physical properties of a cellulose or polysaccharide matrix column and to
 CC modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre.
 CC The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and
 CC has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDs from non-thermophilic organisms

XX Sequence 562 AA;

Query Match 99.4%; Score 1975; DB 3; Length 562;
 Best Local Similarity 99.4%; Pred. No. 7.6e-178;
 Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVWVHGLWSRDYRSMIDQIKSLGYNT 60
 DB 42 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVWVHGLWSRDYRSMIDQIKSLGYNT 101
 QY 61 IRLPYSDDILKPGTMPNSINFQWMDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
 DB 102 IRLPYSDDILKPGTMPNSINFQWMDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 161
 QY 121 SGQSALWYTSVSEATWISDLQALQRYKGNPTVWGFDLHNEPHDPACWCGGSPSIDWRL 180
 DB 162 SGQSALWYTSVSEATWISDLQALQRYKGNPTVWGFDLHNEPHDPACWCGGSPSIDWRL 221
 QY 181 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNQGAGQYPVVLNVNRLVYSAHDY 240
 DB 222 AAEAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNQGAGQYPVVLNVNRLVYSAHDY 281
 QY 241 ATSVYPTWFSDFPTFPNNMGVWYLFNQNTAPVWLGEFGTTLQSTTDQWLKTLV 300
 DB 282 ATSVYPTWFSDFPTFPNNMGVWYLFNQNTAPVWLGEFGTTLQSTTDQWLKTLV 341
 QY 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTVTKDGYLAPIKSSIFDPV 358
 DB 342 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTVTKDGYLAPIKSSIFDPV 399

RESULT 6

AAB48788
 ID AAB48788 standard; protein; 521 AA.

XX AAB48788;

DT 09-MAR-2001 (first entry)

DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.

KW E1 endoglucanase; glycosyl hydrolase; insoluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation; mutant; muten.

OS Acidothermus cellulolyticus.

OS Synthetic.

PN WO200070031-A1.

PD 23-NOV-2000.

PF 19-MAY-2000; 2000WO-US013971.

PR 19-MAY-1999; 99US-0134925P.

PA (MIDE) MIDWEST RES INST.

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;

XX WPI; 2001-061226/07.

XX Preparation of glycosyl hydrolase with an increased catalytic activity on
 XX insoluble substrate.

PS Claim 5; Page 27-29; 30pp; English.

CC The invention relates to a method for making glycosyl hydrolase mutants
 CC with increased catalytic activity with either insoluble or soluble
 CC cellulose substrates relative to the wild-type enzyme. The method for
 CC making a glycosyl hydrolase with increased insoluble substrate catalytic

```

XX 15-JUL-1994; 94US-00276213.
XX (MIDE ) MIDWEST RES INST.
XX Thomas SR, Laymon RA, Himmel ME;
XX WPI; 1996-105843/11.
XX N-PSDB; AAT12337.
XX New isolated DNA encoding endo:glucanase - obtd from Acidothermus
XX cellulolyticus, used for prodn of the enzyme for use in cellulose
XX hydrolysis.
XX Claim 1; Page 22; 34pp; English.
XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful for
XX hydrolysing cellulosic biomass to sugars for simultaneous or subsequent
XX fermentation to ethanol. It shows optimal activity at 83 deg C. The amino
XX acid sequence of the mature enzyme was deduced from an E1 endoglucanase
XX gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides
XX were also identified (see also AAR89928-29). Cloning of this gene allows
XX large-scale, low-cost prodn. of recombinant E1 endoglucanase, using pref.
XX Saccharomyces, Zymomonas or E.coli hosts
XX
XX Sequence 521 AA;
XX
XX Query Match 99.6%; Score 1979; DB 2; Length 521;
XX Best Local Similarity 99.7%; Pred. No. 2.8e-178; Indels 0; Gaps 0;
XX Matches 357; Conservative 0; Mismatches 1;
XX
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
QY 61 IRLPYSDIILKPGTWPNSINFRQWODLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 61 IRLPYSDIILKPGTWPNSINFRQWODLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWGNGLOAGQYPPVNLVNPRLVYSAHDY 240
DB 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWGNGLOAGQYPPVNLVNPRLVYSAHDY 240
QY 241 ATSVYPTQWTSDFTPNNMFGIWNKNWGYLFNQNIAPVWLGEGFTTLQSTTDTQTLKTLV 300
DB 241 ATSVYPTQWTSDFTPNNMFGIWNKNWGYLFNQNIAPVWLGEGFTTLQSTTDTQTLKTLV 300
QY 301 QYLRPTAQYGADSFQWTFWNNPDSGDTGILKDDQWQTVTVKDGXLYAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWNNPDSGDTGILKDDQWQTVTVKDGXLYAPIKSSIFDPV 358

RESULT 4
ID AAU79549 standard; protein; 562 AA.
XX
XX AAU79549;
XX
XX 24-SEP-2002 (first entry)
XX
XX A. cellulolyticus cellulase E1 beta-1,4-endoglucanase precursor.
XX
XX Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;
XX fermentable sugar; ethanol; fermentation; silage; feed; fuel;
XX industrial chemical; biodegradation; chloroaromatic;
XX environmental pollutant; E1 beta-1,4-endoglucanase; el.
XX
XX Acidothermus cellulolyticus.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1. 41
FT /label= Leader_sequence
FT Protein 42. 562
FT /label= Mature_cellulase
XX WO200234926-A2.
XX PN
XX PD 02-MAY-2002.
XX
XX PF 18-OCT-2001; 2001WO-US032538.
XX PR 20-OCT-2000; 2000US-0242408P.
XX (UNMS ) UNIV MICHIGAN STATE.
XX
XX Sticklen MB, Dale BE, Magbool S;
XX WPI; 2002-489947/52.
XX N-PSDB; ABK86729.
XX
XX Producing transgenic plants which after harvest degrade lignin and
XX cellulose to fermentable sugars, by mating transgenic plant comprising
XX DNA encoding cellulase with transgenic plant comprising DNA encoding
XX ligninase.
XX
XX Disclosure; Page 96-99; 126pp; English.
XX
XX The invention discloses the production of a transgenic plant which
XX degrades lignocellulose when the plant is ground. It comprises the
XX production of the transgenic plant including cellulase and ligninase by
XX mating a transgenic plant, containing a DNA encoding a cellulase, and a
XX transgenic plant, containing a DNA encoding a ligninase, where both genes
XX are operably linked to a nucleotide sequence encoding a signal peptide
XX which targets the fusion protein to an organelle of the plant,
XX particularly chloroplasts. The method is useful for producing a
XX transgenic plant (e.g. maize) which degrades lignocellulose when the
XX plant is ground to produce a plant material. This material is useful for
XX converting lignocellulose, in a plant material, to fermentable sugars
XX which are then fermented to ethanol. The transgenic plants also provide a
XX plentiful and inexpensive source of fungal or bacterial cellulases and
XX ligninases which can be used in the production of ethanol. They can also
XX be used for pre-treating silage to increase the energy value of
XX lignocellulosic feeds for cows and other ruminant animals, pre-treating
XX lignocellulosic biomass for fermentative conversion to fuels and
XX industrial chemicals, and biodegradation of chloroaromatic environmental
XX pollutants. The protein sequence presented is the A. cellulolyticus
XX cellulase E1 beta-1,4-endoglucanase precursor
XX
XX Sequence 562 AA;
XX
XX Query Match 99.6%; Score 1979; DB 5; Length 562;
XX Best Local Similarity 99.7%; Pred. No. 3.2e-178; Indels 0; Gaps 0;
XX Matches 357; Conservative 0; Mismatches 1;
XX
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 60
DB 42 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNT 101
QY 61 IRLPYSDIILKPGTWPNSINFRQWODLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 102 IRLPYSDIILKPGTWPNSINFRQWODLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 161
QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 162 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 221
QY 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWGNGLOAGQYPPVNLVNPRLVYSAHDY 240
DB 222 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWGNGLOAGQYPPVNLVNPRLVYSAHDY 281
QY 241 ATSVYPTQWTSDFTPNNMFGIWNKNWGYLFNQNIAPVWLGEGFTTLQSTTDTQTLKTLV 300

```


CC (AA848788) has improved activity with insoluble substrates, and the W42R
CC (AA848786) and Y82R (AA848787) mutants have improved activity with
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase Y82R mutant
XX
SQ Sequence 521 AA;
Query Match 100.0%; Score 1986; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 6.2e-179;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDYRSMDOIKSLGYNT 60
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDYRSMDOIKSLGYNT 60
QY 61 IRLPYSDDLKPGTTPNSINFRMNQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTTPNSINFRMNQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SQGSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQGSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AARAGNAVLNVNPNLLIFVEGVQSYNGDSYWGGLQAGQYFVVLNPNRLVYSAHDY 240
DB 181 AARAGNAVLNVNPNLLIFVEGVQSYNGDSYWGGLQAGQYFVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTQWFSDETFPNNMPCINWKNWGYLFNQNIAPVNLGEGFTTQSTTDTQWLKTLV 300
DB 241 ATSVYPTQWFSDETFPNNMPCINWKNWGYLFNQNIAPVNLGEGFTTQSTTDTQWLKTLV 300
QY 301 QYLPTAQYADSFQWTFWNPDSGDTGGILKDDQWQTVDTVKDGLAPIKSSIFDPV 358
DB 301 QYLPTAQYADSFQWTFWNPDSGDTGGILKDDQWQTVDTVKDGLAPIKSSIFDPV 358
RESULT 2
AAW39262
ID AAW39262 standard; protein; 358 AA.
XX
AC AAW39262;
XX
DT 14-MAY-1998 (first entry)
XX
DE A. cellulolyticus E1-CAT translated region.
XX
KW E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;
KW truncated; cellulose hydrolysis; biomass conversion.
XX
OS Acidothermus cellulolyticus.
XX
XX US5712142-A.
XX
PD 27-JAN-1998.
XX
XX 22-FEB-1996; 96US-00604913.
XX
XX 26-SEP-1989; 89US-00412434.
XX
PR 27-JAN-1992; 92US-00826089.
XX
PR 21-SEP-1993; 93US-00125115.
XX
PR 15-JUL-1994; 94US-00276213.
XX
XX (MIDE) MIDWEST RES INST.
XX
XX Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;
XX
XX WPI; 1998-119985/11.
XX
XX N-PSDB; AAV03659.
XX
XX DNA encoding truncated form of E1 endo:glucanase from Acidothermus

PT cellulolytic - and related vectors and transformed cells, expressing
PT only catalytic domain, is used for biomass conversion and has better heat
PT stability than complete enzyme.
XX
PS Claim 2; Fig 5; 19pp; English.
XX
CC This sequence represents a novel Acidothermus cellulolyticus endoglucanase
CC E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the
CC enzyme, without its peptide linker or cellulose binding domain (CBD) and
CC the truncated enzyme expressed by is used for hydrolysis of cellulose
CC (biomass conversion). Compared with full-length E1, the truncated enzyme
CC has better heat stability and higher temperature of maximum activity
XX
SQ Sequence 358 AA;
Query Match 99.6%; Score 1979; DB 2; Length 358;
Best Local Similarity 99.7%; Pred. No. 1.7e-178;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDYRSMDOIKSLGYNT 60
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDYRSMDOIKSLGYNT 60
QY 61 IRLPYSDDLKPGTTPNSINFRMNQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTTPNSINFRMNQDLQGLTSLQWMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SQGSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQGSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AARAGNAVLNVNPNLLIFVEGVQSYNGDSYWGGLQAGQYFVVLNPNRLVYSAHDY 240
DB 181 AARAGNAVLNVNPNLLIFVEGVQSYNGDSYWGGLQAGQYFVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTQWFSDETFPNNMPCINWKNWGYLFNQNIAPVNLGEGFTTQSTTDTQWLKTLV 300
DB 241 ATSVYPTQWFSDETFPNNMPCINWKNWGYLFNQNIAPVNLGEGFTTQSTTDTQWLKTLV 300
QY 301 QYLPTAQYADSFQWTFWNPDSGDTGGILKDDQWQTVDTVKDGLAPIKSSIFDPV 358
DB 301 QYLPTAQYADSFQWTFWNPDSGDTGGILKDDQWQTVDTVKDGLAPIKSSIFDPV 358
RESULT 3
AAW89927
ID AAW89927 standard; protein; 521 AA.
XX
AC AAW89927;
XX
DT 08-OCT-1996 (first entry)
XX
DE A. cellulolyticus E1 endoglucanase.
XX
KW E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.
XX
OS Acidothermus cellulolyticus.
XX
XX Location/Qualifiers
XX Domain 1..363
XX /label= Catalytic-domain
XX Domain 364..417
XX /label= Linker
XX /note= "proline/serine/threonine-rich linker domain
XX common to multi-domain microbial cellulases"
XX Domain 418..521
XX /label= Cellulose-binding_domain
XX
XX WO9602551-A1.
XX
XX 01-FEB-1996.
XX
XX 14-JUL-1995; 95WO-US008868.


```

QY 223 YPVVL-NVNRNLVYSAHDYATSVYQTW-----FSDPTFPNNMFGIWNKNGYLFNQNIAP 277
DB 1261 YFINLGQYQNKVYSGPHDYGLPYQOPWYFPGFTKDTLYND---CWRDNWITIMDNGIAP 1317
QY 278 VMLGEGTTLQSTDDTQTLWKLTVQYLRLPTAQYCADSFQWTFWSMNPDSGDTGGLIKDDWQ 337
DB 1318 LLIGEGWGLDGDNEKWTYLRDII-----ENHIHFTWCYNANSGDTGGLVGYDFS 1371
QY 338 TVDTVKDGYLAP 349
DB 1372 TWDEQYNFLKP 1383

RESULT 14
Q9EYQ0 PRELIMINARY; PRT; 534 AA.
AC Q9EYQ0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cellulase Cel5-N.
GN CELN.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RA Belaich A., Gaudin C., Gal L.;
RT "A gene cluster of cellulases of Clostridium cellulolyticum.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316923; AAG45162.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR02048; EF-hand.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 534 AA; 59671 MW; 980785F0971E5A59 CRC64;

Query Match 29.6%; Score 587.5; DB 2; Length 534;
Best Local Similarity 32.8%; Pred. No. 3.2e-38;
Matches 133; Conservative 60; Mismatches 144; Indels 69; Gaps 13;

QY 6 W-HTSGREILDANNVPRIAGINWFGFTCNVYVHGLSRDYSMLDQIKSLGYNITIRLP 64
DB 38 WLCVGDGIYDMNGREYVLTGANWFGFNCSENVFHG-AMVDVKNILTSVADRIGILLRVP 96
QY 65 YSDIIL-----KPG-----TMNSINFYQMNQDLOQ-----LTSQVMDKIVAYAGOI 107
DB 97 ISTELLSWMTGRENKYSVTSANNPYTVNPFDPDTPDGPKNSWEIFDIIMKYCKEL 156
QY 108 GLRIILDRHRPDCSGQS-----ALWY-----TSSVSEATWISDLQALAQRYKGNPTVGGFD 158
DB 157 GIKVMIDVHSPDANNSGHMFLWYGLETTTAGMTTIDTKWIDTLTLWLAGYKNDITLAI 216
QY 159 LKNEPH-----DPAWCGGSDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSY- 206
DB 217 LKNEPHKRGYTNAPTDMAKNDNTDENNKVAAERCSKEILAVNPKLLIMEGIEQVP 276
QY 207 -----NGDSY-----WGNLQAGQYPVVLNPN-RLVYSAHDYATSV 244
DB 277 KTEGYTFTDPVWVGASGDAAPWGHGWNGLRGVQDYPIDLGLNLSQIVYSPHDYGPSV 336
QY 245 YPQTWF-SDETFNNMFGIWNKNGYLFNQNIAPVLMGEGTTLQSTDDTQTLWKLTVQYL 303
DB 337 YNQSFWDKDFTTQLDDYWDYTDWAYIDDQKIAPLLIGEGWFGMDGAKQKWTLLRDY 396

```

```

QY 304 RPTAQYCADSFQWTFWSMNPDSGDTGGLIKDDWQTVDTVKDGYLAP 349
DB 397 I-----KNRINHFTWCLNPNSGDTGGLIGNDWSWTWDEEYGLLKP 436

RESULT 15
Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tf cel5B, a new endoglucanase encoding
gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;

Query Match 29.0%; Score 574; DB 2; Length 616;
Best Local Similarity 34.0%; Pred. No. 4.6e-37;
Matches 129; Conservative 63; Mismatches 149; Indels 38; Gaps 13;

QY 1 AGGGYW-HTSGREILDANNVPRIAGINWFGFTCNVYVHGLSRDYSMLDQIKSLGYN 59
DB 33 SGTADWLITDGNRIVDSAGNEVLTGANWFGFNTSERMFHGLWAAANIEDITSAAERGIN 92
QY 60 TIRLPYSDDIL---KPGTM-PNSINFYQMNQDLOGLTSLOVMDKIVAYAGIQLRIILDR 115
DB 93 MVRVPISTQLLEWNGQAGSGVNEY-VNPELAGMTLEVFDTWJQICEYGLKWLMDV 151
QY 116 H--RPDCSGQ-SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----166
DB 152 HSAEADNSGHYYPVWYKGDITTFDTAWEWVTERYKNNDTIVAADIKNEPHGKANETPR 211
QY 167 ACWCGGSDPSID-WRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYW 212
DB 212 AKWD-GSTDIDNFKHVCETAGKRIALNPNMNLILCEGIEIYKDGQDWSSTDGRDYISTW 270
QY 213 WGNLQAGQYPVVLNPN-NRNLVYSAHDYATSVYQTFWFSPTFPNNM-PGIWNKNGYL 270
DB 271 WGNLRGVADHPVDLGAHQDQLVYSPHDYGSVFEQWFEQWNRQTLTLDVWRPNWLYI 330
QY 271 FNQNIAPVLMGEGTTLQSTDDTQTLWKLTVQYLRLPTAQYCADSFQWTFWSMNPDSGDTGG 330
DB 331 HEDDIAPILLIGEWGGLDGDNEKWTALRSIID-----EKMHHTFWALNPNSGDTGG 384
QY 331 ILKDDWQTVDTVKDGYLAP 349
DB 385 LLNYDWTWDEAKYAFIKP 403

Search completed: August 2, 2004, 16:36:57
Job time : 37.3333 secs

```

338 TVDTVKGVLAP 349
 554 TVDEKYAIIIVP 565

Db QY

RESULT 12

024820 PRELIMINARY; PRT; 1000 AA.

ID O24820 AC

DT 01-JAN-1998 (TREMELrel. 05, Created)

DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Beta-glucanase.

OS thermophilic anaerobe NA10.

OC Bacteria.

OX NCBI_TaxID=67756;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NA10.

RA Miyake K., Machida Y., Hattori K., Iijima S.;

RT "Characterization of a multi-domain cellulase from an extremely thermophilic anaerobe strain NA10."

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).

EMBL; AB008029; BAA22939.1; -.

DR HSSP; Q06851; INBC.

DR GO; GO:0003779; F:actin binding; IEA.

DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001589; Actbind actnin.

DR InterPro; IPR001956; CBD 3.

DR InterPro; IPR008965; Cellul bind.

DR InterPro; IPR001000; Glyco_hydro_10.

DR Pfam; PF00942; CBM 3; 1.

DR Pfam; PF00150; cellulase; 1.

DR Pfam; PF00331; Glyco_hydro_10; 1.

DR PRINTS; PR00134; GLHYDLASE10.

DR ProDom; PD001947; CBD_3; 1.

DR SMART; SM00633; Glyco_10; 1.

DR PROSITE; PS00019; ACTININ 1; 1.

DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Glycosidase; Hydrolase.

SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 31.5%; Score 623.5; DB 2; Length 1000;

Best Local Similarity 35.5%; Pred. No. 11e-40;

Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTRILRP 64

Db 590 WLYVSNKIVDXDGRVWLTGVNWFYNTGTNVFGVMSCNLSKSLAEIANRGNLLRVP 649

QY 65 YSDIL---KPGTMPN-SINFYQMNODLOGLTSLQVMDKIVAYAGQIGRIILDRH--RP 118

Db 650 ISAEILNWSQGIYKPNINYY-VNPELGKNSLEVFIVVQTKVEVGLKIMLDIHSIKT 708

QY 119 DCSGQ-SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDP-----ACWG 170

Db 709 DAMGHYIPVWYDEKFTPEFYKACWITRYKNDTIIAFDLKNEPHKPNQDTTFKWD 768

QY 171 CGDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYNGD-----SYWGGNL 217

Db 769 NSTDINNWKYAAETCAKRIILNPNLLIIEGIEAYPKDDVTWTSKSSDYSTWGGNL 828

QY 218 QGAGQYPVVL-NVPEFLVYSAHDYATSYVPOTWPSDPTFENN--MPGLWKNWGLFNQ 274

Db 829 RGVKYPINLKGKYNKVYSPHDYGSVYQQPWFY-PGFTKESLQDCWRNWAYIMEEN 887

QY 275 IAPVWLGEFGTTILQSTTDDTKLTVQVLRPTAQYGADSFQWTFMSWNPDSGDTGGILKD 334

Db 888 IAPLLIGEWGGLDGDADNEKWKYLRDVII-----ENHIHTTWCNFGNSGDTGGVLGY 941

QY 335 DMQTVDTVKGVLAP 349

Db 942 DFTTWDEKKYGFLLKP 956

RESULT 13

Q9X3P6 PRELIMINARY; PRT; 1426 AA.

ID Q9X3P6 AC

DT 01-NOV-1999 (TREMELrel. 12, Created)

DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE CelB.

GN CELB.

OS Caldicellulosiruptor sp. Tok7B.1.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;

OC Caldicellulosiruptor.

OX NCBI_TaxID=80339;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Tok7B.1;

RX MEDLINE=20171169; PubMed=10706665;

RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P., Bergquist P.L.;

RT "Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1.";

RL Curr. Microbiol. 40:333-340(2000).

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).

EMBL; AF078737; AAD30364.1; -.

DR HSSP; Q06851; INBC.

DR GO; GO:0003779; F:actin binding; IEA.

DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001589; Actbind actnin.

DR InterPro; IPR001956; CBD 3.

DR InterPro; IPR008965; Cellul bind.

DR InterPro; IPR001000; Glyco_hydro_10.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00942; CBM 3; 3.

DR Pfam; PF00150; cellulase; 1.

DR Pfam; PF00331; Glyco_hydro_10; 1.

DR PRINTS; PR00134; GLHYDLASE10.

DR ProDom; PD001947; CBD_3; 3.

DR SMART; SM00633; Glyco_10; 1.

DR PROSITE; PS00019; ACTININ 1; 1.

DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Glycosidase; Hydrolase.

SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 31.1%; Score 617; DB 2; Length 1426;

Best Local Similarity 35.5%; Pred. No. 5.7e-40;

Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTRILRPYSDDI 69

Db 1022 GNKIVDKDGKPVWLTGVNWFYNTGTNVFGVMSCNLSKSLAEIANRGNLLRVPISAEI 1081

QY 70 L---KPGTMPN-SINFYQMNODLOGLTSLQVMDKIVAYAGQIGRIILDRH--RPDCSGQ 123

Db 1082 ILNWSKGYKPNINYY-VNPELGKLSLEVFDFVVKCKEVGLKIMLDIHSAKTDAMGH 1140

QY 124 -SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDP-----ACWCGDPS 175

Db 1141 IYVWYTDITITPEDIYKACEWITERYKNDTIIAFDLKNEPHKPNQDVTWTSKSSDYSTWGGNL 1200

QY 176 IDWRLAERAGNAVLSVNPILLIFVEGVQSYNGD-----SYWGGNLQAGQ 222

Db 1201 NNWKYAAETCAKRIILNPNLLIIEGIEAYPKDDVTWTSKSSDYSTWGGNLRGVKK 1260

338 TVDTVKGVLAP 349
 554 TVDEKYAIIIVP 565

Db QY

RESULT 12

024820 PRELIMINARY; PRT; 1000 AA.

ID O24820
 AC O24820;
 DT 01-JAN-1998 (TREMELrel. 05, Created)
 DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Beta-glucanase.
 OS thermophilic anaerobe NA10.
 OC Bacteria.
 OX NCBI_TaxID=67756;
 RN [1]
 RC STRAIN=NA10.
 RP SEQUENCE FROM N.A.
 RA Miyake K., Machida Y., Hattori K., Iijima S.;
 RT "Characterization of a multi-domain cellulase from an extremely
 thermophilic anaerobe strain NA10."
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
 HYDROLASES).
 DR EMBL; AB008029; BAA22939.1; -.
 DR HSSP; Q06851; INBC.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001589; Actbind actnin.
 DR InterPro; IPR001956; CBD 3.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001000; Glyco_hydro_10.
 DR Pfam; PF00942; CBM 3; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00331; Glyco_hydro_10; 1.
 DR PRINTS; PR00134; GLHYDLASE10.
 DR ProDom; PD001947; CBD_3; 1.
 DR SMART; SM00633; Glyco_10; 1.
 DR PROSITE; PS00019; ACTININ 1; 1.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Glycosidase; Hydrolase.
 SQ SEQUENCE 1000 AA; 113265 MW; 89659A56A752C6B CRC64;

Query Match 31.5%; Score 623.5; DB 2; Length 1000;
 Best Local Similarity 35.5%; Pred. No. 11e-40;
 Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPRVIRAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTRILRP 64
 DB 590 WLYVSNKIVDXDGRVWLTGVNWFYNTGTNVFGVMSCNLSKSLAEIANRGNLLRVP 649
 QY 65 YSDIL---KPGTNPEN-SINFYOMNODLOGLTSLQVMDKIVAYAGQIGRIILDRH--RP 118
 DB 650 ISAEILINWSQGIYKPNINYY-VNPELGKNSLEVFIVVQTKVEVGLKIMLDIHSIKT 708
 QY 119 DCSGQ-SALWYTSVSEATWISDLQALARYKGNPTVVGFDLHNEPHDP-----ACWG 170
 DB 709 DAMGHYPVWYDEKFTPEFYKACWITRYKNDTTIAFLDKNEPHKPNQDTTFKWD 768
 QY 171 CGDPSIDWRLAAERAGNAVLSVNPILLIFVEGVQSYNGD-----SYWGGNL 217
 DB 769 NSTDINNWKYAAETCAKRIINLPNLLIVIEGIEAYPKDDVTWTSKSSDYSTYTWGGNL 828
 QY 218 QGAGQYPVVL-NVPEFLVYSAHDYATSYVPTWPSDPTFENN--MPGLWKNWGLFNQ 274
 DB 829 RGVKYPINLKGYNKVYSPHDYGSVYQQPWFY-PGFTKESLQDCWRNWAYIMEEN 887
 QY 275 IAPVWLGEFGTTILQSTTDDTKLTVQVLRPTAQYGADSFQWTFMWNPDSDGTGGILKD 334

```
QY 338 TVDTKVDGYLAP 349
Db 377 TIDTKLALVQP 388

RESULT 10
Q87AH4 PRELIMINARY; PRT; 614 AA.
AC Q87AH4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN ENGXA OR P1851.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=193190;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=22421331; PubMed=12533478;
RX Van Slyke M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.W., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sana J.A.D.,
RA de Souza A.A., Truffi D., Takumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026 (2003).
DR EMBL; AE012560; AAC29683.1; -
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002952; Egshe11.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Complete proteome.
KW SEQUENCE 614 AA; 62041 MW; 29EC67E42F4BC3F7 CRC64;

Query Match 32.6%; Score 645.5; DB 16; Length 614;
Best Local Similarity 39.9%; Pred. No. 9.6e-43;
Matches 141; Conservative 58; Mismatches 125; Indels 29; Gaps 12;

QY 5 YHSTGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNITRLP 64
Db 26 YSI:SHKGVVDKGN-QIQLEGVNWFGFETGHDVNVGLWLNWKEFITQLQGMGFNAIRLP 84

QY 65 YSDDLKPGTMTMNSINFYQNDLQGLTSLQVMDKIVAYAGQIGLRIILDRHPCSGOS 124
Db 85 FCPANLNSNTSPSSID-YSENPLQGLSLQILDVKVKLSDRMYTVLLDHRHPCDSATS 143

QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDACWCGGPPSIDRWLAER 184
Db 144 ELWYTSYSEKQIDDLRFVHARYANVGHVIGLEVNNEPHRTTWGDPKTDWNTAVEH 203

QY 185 AGNAVLSVNPILLIFVEGVOSYNGDS-----YWGNGNLQAGQYPPVNLVP-NRLVY 235
Db 204 AAAAILEAAPKWLIGVEGI-----GENPICSTTICHFNGENLEPMDCPT--LKPVPADHLL 257

us-09-997-504a-12.rspt

QY 236 SAHDYATSVYPTQWFSDFTPNNMPGIVNKNWGLFNQNTAPVWLGEFGTTL--QSTTDQ 293
Db 258 MPHVGPDVYVQYFNSPDPNNMAIWDKHFHFAKAGYA-MAIGFEGYKGEGRDI 316

QY 294 TWLKTLYQYLRPTAQYG-ADSFQWTFWSNPDSDGTGGILKDDQWTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAP---YWAANQNSVDIGGMVGNWNTTPRDDKVK 363

RESULT 11
Q9L3J2 PRELIMINARY; PRT; 660 AA.
AC Q9L3J2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-cellobiohydrolase.
GN CELO.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=F7;
RA Zverlov V.V., Schwarz W.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ275975; CAB76938.1; -
DR HSSP; P54583; 1BCE.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR Hydrolyase.
KW SEQUENCE 660 AA; 75247 MW; 04A042002C288CFC CRC64;

Query Match 32.0%; Score 634.5; DB 2; Length 660;
Best Local Similarity 36.0%; Pred. No. 7.9e-42;
Matches 134; Conservative 66; Mismatches 137; Indels 35; Gaps 10;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCN-VYVHGLRSRDYRSMLDQIKSLGYNITRL 63
Db 201 WLHVEGNLIKDAQGNTVYLTGINWFGFETGANGFHGNKCNLEDSLDLMKLGFNILRI 260

QY 64 PYSDDL---KPGTMTMNS--INFYQNMNDLQGLTSLQVMDKIVAYAGQIGLRIILDRH-R 117
Db 261 PISAEIILQWNGRVERVTSFVNTYE-NPRLDGLSSILDYTNHMKNGKMAIDMESS 319

QY 118 PDCSGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH-----DPAWGC 171
Db 320 TKDSYQENLWNTKDIITWEEFTEAWKWIYKDDDTVIADVLDKNEPHGKYSGNIAKWD 379

QY 172 GDPSTDLWLAERAGNAVLSVNPILLIFVEGVOSYNGDSY-----WVGNLQ 218
Db 380 SNDPNNWKRRAEIIABEILAINPNLLIVVEGEVAYPMEGDYTCNGEFTYCNWVGNNLR 439

QY 219 GAGQYPPVNLVPNRLVLSAHDYATSVYPTQWFSDFTPNNMPGIVNKNWGLFNQNTAP 277
Db 440 GVADHPVVISAPDKLVYSHVDYGPDIYMQPKFKDFDINTLYEECYPNWYIYEQNIAP 499

QY 278 VWLGEFGTTLQSTDDQWTKLVQYLRPTAQYGADSFQWTFWSNPDSDGTGGILKDDQ 337
Db 500 MLIGEWGKGLINENNRKWLCELATFI-----ASKKLHHTFWFNPNSADTGGGLNLEDWK 553
```

OC Xanthomonadaceae; Xylella.
 ON NCBI_TaxID=2371;
 RN [1]
 RP
 RC SEQUENCE FROM N.A.
 RX STRAIN=9a5c;
 RA MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu P.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.B., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Marques M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki A.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Teuhako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RA "The genome sequence of the plant pathogen Xylella fastidiosa";
 RT Nature 406:151-159(2000).
 DR EMBL; AF003921; AF83628.1; -.
 DR PIR; E82759; E82759.
 DR HSSP; P54583; 1ECE.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001919; Bac_celose-bind.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00553; CBM 2; 1.
 DR SMART; SM00637; CBD II; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Complete proteome.
 KW SEQUENCE 592 AA; 59967 MW; 9846DA4EA3B5C89E CRC64;
 SQ
 Query Match 34.5%; Score 683.5; DB 16; Length 592;
 Best Local Similarity 42.6%; Pred. No. 8.7e-46;
 Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;
 QY 5 YWHTSGREILDANNVPVRIAGINWFGFTTCNVVHGLRSRDRYRMLDQIKSLGYNITRLP 64
 DB 26 YSTSHGKVIDDKGN-QIOLKGSWFGFTTHVHGLTRNWKFEFTIQISMGLENVRLP 84
 QY 65 YSDDLKPGTMTNSINFYQMNODLGLTSLQWMDKIVAYAGIQLRIILDRHRDPCSGQS 124
 DB 85 FCPASLNSNTSPSSID-YNRNPDLOGLSLQMDKVKELSDRIYVLMHHTFDCAIS 143
 QY 125 ALMYTSSVSSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWGCGDPSIDNRLAER 184
 DB 144 ELWYTDYSSEKQWIDDLREVAHYANVPVIGLDVKNPEPHGRATWGTGDPKTDWNTAVEH 203
 QY 185 AGNVLNPNLLIIFVEGVQSYNGDS----YWGGLAGAGQYFVLNVP-NRLVYSAHD 239
 DB 204 AAAAILAEPKVLIGVEGIGENPSCSSTIGHFWGLENLEFMDCTP--LKVPADHLLIPHV 261
 QY 240 YATSVYFQWFSDDPTFPNNMPCINWKNWGLYFNQNIAPVWLGFEFTIL--QSTTDITWLK 297
 DB 262 YGPDVTVQYFNSPDPFNMAALWKHFGHAKAGYA-NAIGETFGKYGEGRDPAWQON 320

OC Xanthomonadaceae; Xylella.
 ON NCBI_TaxID=2371;
 RN [1]
 RP
 RC SEQUENCE FROM N.A.
 RX STRAIN=9a5C;
 RA MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu P.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.B., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Marques M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki A.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Teuhako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RA "The genome sequence of the plant pathogen Xylella fastidiosa";
 RT Nature 406:151-159(2000).
 DR EMBL; AF003921; AF83628.1; -.
 DR PIR; E82759; E82759.
 DR HSSP; P54583; 1ECE.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001919; Bac_celose-bind.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00553; CBM 2; 1.
 DR SMART; SM00637; CBD II; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Complete proteome.
 KW SEQUENCE 592 AA; 59967 MW; 9846DA4EA3B5C89E CRC64;
 SQ
 Query Match 34.5%; Score 683.5; DB 16; Length 592;
 Best Local Similarity 42.6%; Pred. No. 8.7e-46;
 Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;
 QY 5 YWHTSGREILDANNVPVRIAGINWFGFTTCNVVHGLRSRDRYRMLDQIKSLGYNITRLP 64
 DB 26 YSTSHGKVIDDKGN-QIOLKGSWFGFTTHVHGLTRNWKFEFTIQISMGELNAVRLP 84
 QY 65 YSDIILKPGTMTNSINFYQMNODLGLTSLQWMDKIVAYAGIQLRIILDRHRDPCSGQS 124
 DB 85 FCPASLNSNTSPSSID-YNRNPDLOGLSLQMDKVKELSDRIYVLMVHHTFDCAIS 143
 QY 125 ALMYTSSVSSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWGCGDPSIDNRLAER 184
 DB 144 ELWYTDYSSEKQWIDDLREVAHYANVPVIGLDVKNPEPHGRATWGTGDPKTDWNTAVEH 203
 QY 185 AGNVLVSNPNLLI FVEGVQSYNGDS----YWGGLAGAGQYFVLNVP-NRLVYSAHD 239
 DB 204 AAAAILAEPKVLIGVEGIGENPSCSSTIGHFWGLENLEFMDCTP--LKVPADHLLIIPHV 261
 QY 240 YATSVYFQWFSDDPTFPNNMPCINWKNWGLYFNQNIAPVWLGFEFTIL--QSTTDITWLK 297
 DB 262 YGPDVTVQYFNSPDPFNMAALWKHFGHAKAGYA-NAIGETFGKYGEGRDPAWQON 320

```
Db 262 YGPDVFVQSYFNDNFNNPAINRHFQFAGSH--ALLLGEFGKYGEGDARDKQWQD 319
QY 298 TLVQYLRLPTAQYGADSFQWTFWNSWNPDSGDTGGLKDDWQTV 339
Db 320 ALVKYLR---SKGIN---EGFYWSWNPNSGDTGGLRDDWTSV 356

RESULT 6
O58925 PRELIMINARY; PRT; 458 AA.
AC O58925;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 458AA long hypothetical endo-1,4-beta-glucanase.
GN PH1171.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawaiyayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamanoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyana A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000005; BAA30271.1; -.
DR PIR; E71059; E71059.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco hydro_5.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW Complete proteome.
SQ SEQUENCE 458 AA; 51930 MW; CAE48AD43A8EE654 CRC64;

Query Match 37.9%; Score 751; DB 17; Length 458;
Best Local Similarity 45.8%; Pred. No. 2.6e-51;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNTIRLPYSD 67
Db 54 TSGEE-----TPIHLFGVNWFGFETPNVHGLWKNWEDMLQIKSLGFNAIRLPFCT 107
QY 68 DILKPGTMPSINFYQNDQGLTSLOWDKIVAGQIGLRIILDRHPCSGOSALW 127
Db 108 ESKVPGTQPTGID-YSKNPLRLGLDSLIQMEIKIKKAGDLGIFVLLDYHRICTHIEPLW 166
QY 128 YTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSSEDFINTWIEVAKFGKXVNVIGADLNKNEPHSVTSPPAAVTDGTGATWGNPFA 226
QY 176 IDWRLAERAGNAVLSNPNLLIFVEGVQSYN-----GDSYWGNGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVPAPHLIFVEGTQFTNPKTDSYKWKGNWNGNLMAVKDYPV- 285
QY 227 LNVP-NRLVVSADHYATSVVPQTWFSDEPT-FPNMPPGIMNKNWGYLFNQNIAPVWLGEFG 284
Db 286 -NLPRNLKLVSPHYGPDVNVQYFGPAKGFPPNLDPIWTHHGYKVLGLYSVWVIGFG 344
QY 285 TTL---QSTTDQTKLTKLVQYLRPTAQYGADSFQWTFWNSWNPDSGDTGGLKDDWQTV 339
Db 345 GKYGHGDPDRDVIWQNKLVDMW---IENKFDFF---YWSWNPDSGDTGGILQDDWTI 397
```

```
RESULT 7
Q9V052 PRELIMINARY; PRT; 514 AA.
AC Q9V052;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase.
GN CEL3-LIKE OR PYRAB09400 OR PAB0632.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; CAB49854.1; -.
DR PIR; E75142; E75142.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 514 AA; 59980 MW; CF80CAB1525337C5 CRC64;

Query Match 35.0%; Score 694.5; DB 17; Length 514;
Best Local Similarity 41.9%; Pred. No. 9.6e-47;
Matches 149; Conservative 57; Mismatches 121; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60
Db 30 YTAENGIIFVQNVTTGEEKKPLYLHGVSFWGFELKHVYVGLDKENWKDILKDKVRLGFNA 89
QY 61 IRLPYSDDILKPGTMN--SINFYQMDLQGLTSLOWDKIVAGQIGLRIILDRHP 118
Db 90 IRLPFCSEIRPDRPSPERIN-YELNPDNLKLTSLIMEIKLIETANSIGLYLLDYHRI 148
QY 119 DCSQSALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPAWCGDPSIDW 178
Db 149 GCEETIEPLWYTYNSEEQYIKDWIFLAKRFKYPNVIGADIKNEPHGEAGTGDGR-DF 207
QY 179 RLAAERAGNAVLSNPNLLIFVEGVQ-----SYNGDSYV--WGNLQAGQYPVV 227
Db 208 RLFAEKVGREILKVPAPHLIFVEGTQFTNPKTDSYKWKGNWNGNLMAVKDYPVRL 267
QY 228 NVPNRLVVSADHYATSVVPQTWFSDEPTFPNMPGIMNKNWGYLFNQNIAPVWLGEFGTTL 287
Db 268 -PRGVVYSPHYGPDVNVQYFGPAKGFPPNLDPIWTHHGYKVLGLYSVWVIGFG 325
QY 288 QSTTDQTKLTKLVQYLRPTAQYGADSFQWTFWNSWNPDSGDTGGILKDDWQTVDK 343
Db 326 EG-LDKVWQDAFVKWLKIKKIY-----NFFYCLNPDSGDTGGIFLDWKTNNWEK 375

RESULT 8
Q9PF60 PRELIMINARY; PRT; 592 AA.
AC Q9PF60;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN XP0818.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
```

Db 165 SELWYTSYPSRWSIDRMLAKRYASDPTVIGVDLHNEPHGAATWGTGAATDWRAAA 224
 Qy 184 RAGNAVLSVNPMLLI FVEGV-QSYNGDSYWGNGLQAGAGQPVVLANVPNELVYSADHYAT 242
 Db 225 RGGNAVLAENPKLLVLEGDHQAADGTGTWGGALDSAAATASVRLTVANRVVTSPHDYP 284
 Qy 243 SVYPTQWFSPTFFNNPFGIWNKNGVLFNQNIAPVWLGEFGTTLQSTDTQTLKTLVQY 302
 Db 285 TIYQGFWSASNYTNLPGLIWDHAGYLAKKDIAPVLVGEFGTKLEFASDKQWLNTLVGY 344
 Qy 303 LRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPI 350
 Db 345 LSST---GISS---SFAFNPDSDGTGGIVKSDWVTPEQAKLDALAPI 386

RESULT 4
 Q9EVR6 PRELIMINARY; PRT; 518 AA.
 AC Q9EVR6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-OCT-2002 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Extracellular endoglucanase (ENGXCA protein) precursor.
 GN ENGXCA.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Schroeter K., Puchler A., Becker A.;
 RA "engXCA major extracellular endoglucanase";
 RL Submitted (DRC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ304415; CAC18529.1; --
 DR HSSP; P54583; 18CE.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR SIGNAL.
 FT SIGNAL.
 FT CHAIN.
 SQ SEQUENCE 518 AA; 55588 MW; 9A0FF678B3BF712C CRC64;

Query Match 40.0%; Score 792.5; DB 2; Length 518;
 Best Local Similarity 46.9%; Pred. No. 1.6e-54;
 Matches 161; Conservative 59; Mismatches 104; Indels 19; Gaps 9;

Qy 5 YWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNTRILP 64
 Db 26 YSINNSRQIVDSGKVLQKGVNFGFETGNVHVMHGLWARNKEMIVQMGLGFNAVRLP 85
 Qy 65 YSDILKPGTWPNSINFQYNQDQLGTLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 124
 Db 86 FCPATLRSDTWPAID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHTPDCAIS 144
 Qy 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
 Db 145 ELWYTSYTEAQWLADLRFVANYRKNVPYVLGLDLKNEPHGAATWGTGNAATDWNKAAER 204
 Qy 185 AGNAVLSVNPMLLI FVEGVQ-----SYNGDSYWGNGLQAGAGQPVVLANVP-NELVYSAH 238
 Db 205 GSAVLAVAPKWLIAVEGITDNPVCSYNG-GIFWGNLQPLACTP--LNPANRLLLAPH 261
 Qy 239 DYATSVYPTQWFSPTFFNNPFGIWNKNGVLFNQNIAPVWLGEFGTTL--QSTDTQTL 296
 Db 262 VYGPDPVFQSYFNSFNPNMFAIWERHFGQFAGTH--ALLGFGGKYGEGDARDKTMQ 319
 Qy 297 KTLVQYLRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTV 339

Db 320 DALVKYLR---SKGIN--QGFYWSWNPNSGDTGGILRDDWTSV 357
 RESULT 5
 Q8PPS3 PRELIMINARY; PRT; 474 AA.
 AC Q8PPS3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cellulase.
 GN ENGXCA OR XAC0612.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M.,
 RA Locali B.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira I.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RT Nature 417:458-463(2002).
 DR EMBL; AB011689; AAM3501.1; --
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001919; Bac celose-bind.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00553; CBM 2; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Complete proteome.
 SQ SEQUENCE 474 AA; 51294 MW; 379884D63CASA31B CRC64;

Query Match 39.9%; Score 791.5; DB 16; Length 474;
 Best Local Similarity 46.2%; Pred. No. 1.7e-54;
 Matches 158; Conservative 61; Mismatches 104; Indels 19; Gaps 9;

Qy 6 WHTSGREILDANNVPRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNTRILP 65
 Db 26 YSVSNKKIVDDNGNVQLKGVNFGFETGNVHVMHGLWARNKEMINQMGLGFNAVRLP 85
 Qy 66 SDDILKPGTWPNSINFQYNQDQLGTLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGOSA 125
 Db 86 CPATLRSDTWPAID-YSRNADLQGLTSLQILDKVINEFNARGMYVLLDHTPDCAISE 144
 Qy 126 LWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 185
 Db 145 LAYTSYSEAOQLDRLRFVANYRKNVPYVLGLDLKNEPHGAATWGTGNAATDWNKAAER 204
 Qy 186 GNAVLSVNPMLLI FVEGVQ-----SYNGDSYWGNGLQAGAGQPVVLANVP-NELVYSAD 239
 Db 205 SAAVLAVAPKWLIAVEGITDNPVCSYNG-GIFWGNLQPLACTP--LNPANRLLLAPH 261
 Qy 240 YATSVYPTQWFSPTFFNNPFGIWNKNGVLFNQNIAPVWLGEFGTTL--QSTDTQTLK 297


```

QY 124 SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 156 SELWYTAGKIVTASGAPYTIIRGIAWFMETSSCAPHGLDTITLACGMQHKKQMGFTTVRL 129
QY 184 RAGNAVLSVNPVLLIFVEGV---SYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHDY 240
Db 216 RAGNAVLSVNPVLLIFVEGV---SYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHDY 275
QY 241 ARSVPTWFSPTFPNNPFGIWNKMGYLFNQTAPVWLGFBG--TTLOSTTDTWLT 298
Db 276 GPCVSSQPFNDSTFNSNLPALWDQWGYISKQNTAPVLVGFGRNVLDLSSPEKQWNA 335
QY 299 LVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDMTQVTVKDGVIAPKSSIFPPV 358
Db 336 LVDYI-----GANNLYFYWISLNFNSGDTGGLLDDTTWNRPKQDML----SRIMKPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
AC Q9K5C7;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OG Plasmid pCMI.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Meleuzus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-614 FROM N.A.
RC STRAIN=NCPPB 382;
RX MEDLINE=20331586; PubMed=10875331;
RA Jahr H., Dreier J., Meleuzus D., Bahro R., Eichenlaub R.J.;
RT "The endo-beta-1,4-glucanase Cella of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato.";
RL Mol. Plant Microbe Interact. 13:703-714 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Jahr H.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X62582; CAA44467.2; -.
DR HSP; P54583; IECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Plasmid; Signal.
FT SIGNAL 66 POTENTIAL.
FT CHAIN 67 >745 CELLULASE.
FT NON_TER 745
SQ SEQUENCE 745 AA; 77886 MW; 5B45015EB8D4F27C CRC64;

Query Match 46.1%; Score 913.5; DB 2; Length 745;
Best Local Similarity 50.3%; Pred. No. 6.4e-64;
Matches 175; Conservative 51; Mismatches 111; Indels 11; Gaps 4;

```

```

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSMLDQIKSLGYNTIRL 63
Db 70 GMLHTAGKIVTASGAPYTIIRGIAWFMETSSCAPHGLDTITLACGMQHKKQMGFTTVRL 129
QY 64 PYSDILKPGTNPNSINFYQNMQDLOGLTSLQWMDKIVAYAGQIGRIILDRHRPDCSGQ 123
Db 130 PFSNQCL-----AASGVTVGSADPSLAGLTPLOVMDHVVASAKSAGLDVILDOHRPDSGGQ 185
QY 124 SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 186 SELWYTSQYPSRWISDWRMLAKRYAAEPTVIGVDLHNEPHGAATWGTGAATDWRAAAE 245
QY 184 RAGNAVLSVNPVLLIFVEGV---QSYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHDYAT 242
Db 246 RGNNAVLAENPNLLVLEGIHDEADSGTWWGCALGLVGNAPVRLSVANRVVYSPHDYPS 305
QY 243 SVYPTWFSPTFPNNPFGIWNKMGYLFNQTAPVWLGFBGTTLOSTTDTWLTWLTQVY 302
Db 306 TIYQSWFSASNYPANLPGIWDHAGYLAKKDAPVLVGFGRNVLDLSSPEKQWNA 358
QY 303 LRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDMTQVTVKDGVIAPKSSIFPPV 350
Db 366 LSST---GISS---SFWAFNPNSGDTGIVKSDWVTPQAKLDALAPI 407
QY 366 LSST---GISS---SFWAFNPNSGDTGIVKSDWVTPQAKLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
AC Q9AF65;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cellulase Cella.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Laine M., Haapalainen M., Wahlroos T., Kankare K., Nissinen R.,
RA Kassuwi S., Mettler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain.";
RL Physiol. Mol. Plant Pathol. 0:0-0(2001).
DR EMBL; AY007311; AAK16222.1; -.
DR HSP; P54583; IECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 727 AA; 75811 MW; 95DF9B664873985D CRC64;

Query Match 45.7%; Score 906.5; DB 2; Length 727;
Best Local Similarity 50.3%; Pred. No. 2.2e-63;
Matches 175; Conservative 48; Mismatches 114; Indels 11; Gaps 4;

```

```

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSMLDQIKSLGYNTIRL 63
Db 49 GMLHTAGKIVTASGAPYTIIRGIAWFMETSSCAPHGLDTITLACGMQHKKQMGFTTVRL 108
QY 64 PYSDILKPGTNPNSINFYQNMQDLOGLTSLQWMDKIVAYAGQIGRIILDRHRPDCSGQ 123
Db 109 PFSNQCL-----AASGVTVGSADPSLAGLTPLOVMDHVVASAKSAGLDVILDOHRPDSGGQ 164
QY 124 SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183

```

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 ; Search time 36.3333 Seconds
(without alignments)
3108.867 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982

Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phage:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp rvirus:*
- 16: sp bacteriap:*
- 17: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1044	52.7	397	2 Q8RP23	Q8rp23 paenibacill
2	913.5	46.1	745	2 Q9K5C7	Q9k5c7 claviibacter
3	906.5	45.7	727	2 Q9AF65	Q9af65 claviibacter
4	792.5	40.0	518	2 Q9EV86	Q9ev86 xanthomonas
5	791.5	39.9	474	16 Q8P8S3	Q8p8s3 xanthomonas
6	751	37.9	458	17 Q58325	Q58325 pyrococcus
7	694.5	35.0	514	17 Q9V052	Q9v052 pyrococcus
8	683.5	34.5	592	16 Q9PF60	Q9pf60 xylella fas
9	674	34.0	482	16 Q97KK6	Q97kk6 clostridium
10	645.5	32.6	614	16 Q87AH4	Q87ah4 xylella fas
11	634.5	32.0	660	2 Q9L3J2	Q9l3j2 clostridium
12	623.5	31.5	1000	2 Q24820	Q24820 thermophil
13	617	31.1	1426	2 Q8X3B6	Q8x3b6 caldicellul
14	587.5	29.6	534	2 Q9EYQ0	Q9eyq0 clostridium
15	574	29.0	616	2 Q7X2N2	Q7x2n2 thermomonos
16	509.5	25.7	425	2 Q93Q07	Q93q07 thermus cal

17	498	25.1	341	2 P96310	P96310 anaerocellu
18	485	24.5	574	2 Q8S3V3	Q8s3v3 cellulomona
19	398	20.1	630	2 Q8JY77	Q8jy77 stigmatella
20	328	16.5	565	16 Q9PF68	Q9pf68 xylella fas
21	317.5	16.0	569	16 Q87AG9	Q87ag9 xylella fas
22	281.5	14.2	590	16 Q8P513	Q8p513 xanthomonas
23	267	13.5	508	10 Q9LTM8	Q9ltm8 arabidopsis
24	252.5	12.7	522	10 Q9SAB6	Q9sab6 arabidopsis
25	214	10.8	526	10 Q9LFS2	Q9lfs2 arabidopsis
26	204.5	10.3	488	10 Q9LFE7	Q9lfe7 arabidopsis
27	194.5	9.8	354	16 Q8PEF2	Q8pef2 xanthomonas
28	190.5	9.6	561	10 Q9LTN0	Q9ltn0 arabidopsis
29	189.5	9.6	555	10 Q7XU04	Q7xu04 oryza sativ
30	181.5	9.2	437	3 Q99036	Q99036 trichoderma
31	180.5	9.1	694	2 Q9Z187	Q9z187 bacillus st
32	180	9.1	439	3 Q9Z401	Q9z401 agaricus bi
33	179	9.0	439	3 Q9P893	Q9p893 agaricus bi
34	178.5	9.0	377	3 Q00012	Q00012 aspergillus
35	168.5	8.5	558	10 Q7XU06	Q7xu06 oryza sativ
36	166.5	8.4	357	16 Q8PRD3	Q8prd3 xanthomonas
37	166	8.4	550	10 Q7XU05	Q7xu05 xanthomonas
38	165	8.3	377	16 Q8PRD5	Q8prd5 xanthomonas
39	163	8.2	364	2 Q9F0G8	Q9f0g8 rhizobium m
40	157.5	7.9	363	2 Q07652	Q07652 cellvibrrio
41	155.5	7.8	516	2 Q66185	Q66185 bacillus ci
42	154.5	7.8	722	16 Q97L56	Q97l56 clostridium
43	153	7.7	357	2 Q59441	Q59441 fibrobacter
44	151.5	7.6	356	16 Q9PA12	Q9pal2 xylella fas
45	149.5	7.5	329	16 Q9X274	Q9x274 thermotoga

ALIGNMENTS

RESULT 1

Q8RP23 ID Q8RP23 PRELIMINARY; PRT; 397 AA.

AC Q8RP23; RC STRAIN=KCTC 8948P;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Endo-beta-1,4-glucanase

OS Paenibacillus sp. KCTC8948P

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI_TaxID=109199;

RN [1]

RP SEQUENCE FROM N.A.

RA Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;

RT "Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-expression with Endomycetes fibuliger beta-glucosidase gene in Saccharomyces cerevisiae."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF345984; AAL83749.1; ..

DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00150; cellulase; 1.

DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.

SQ SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;

Query Match 52.7%; Score 1044; DB 2; Length 397;

Best Local Similarity 54.7%; Pred. No. 1.1e-74;

Matches 197; Conservative 44; Mismatches 103; Indels 16; Gaps 5;

QY 4 GYWHTSGRILDANNVPRIAGINWFGTETCNVYVHGRSDYRSMLOIKSLGNTLRL 63

DB 37 GYVHTQGNKIVDETCKEAAFGNLFGLTPTNYTLHLGWSRSMDDMLDQVKKEGYNLRL 96

QY 64 PYSDILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGIQLRIILDRHRPDCSGQ 123

DB 97 PYSNQLFDSSSRADSIDYIK-NPDJLVGLTPTIQIMDKLIEKAGQRIQIILDRHRPDCSGQ 155

This Page Blank (uspto)

Query Match 7.4%; Score 147.5; DB 1; Length 814;
Best Local Similarity 20.7%; Pred. No. 0.00061;
Matches 80; Conservative 61; Mismatches 134; Indels 111; Gaps 19;
6 WHSGRILLDANNVPRIAG-----INWFGPCTCNVYVHGL--RSRDYRSM 49
32 FQVSGTKLLDASGNELVWRGMRDISAIDLKVEIKIGNLGNLTLDATETAWGNPRTTKAM 91
50 LDOIKSLGYNTIRLPYSDDILKPTMPNSINFQMNQDLQGLTSLQVMDKIVAYAGQIGL 109
92 IEKVREMGFNARVPVTDW-THIGPADP-----YKIDB-----AWLNRVEVNVVLDGCM 141
110 RIILDRHPCSGQSALWYTSVSEATWISDLOA-----LAQRYKGNPTV 154
142 YALINLHH-----DNTWIIPTVANEORSKEKLVKWEQIATRFKDYDDH 185
155 VGFOLNHPHDPACWCGDPSIDWRLAERAGNAVLSVNPENLLIFVEGVQSYNGDSYWMG 214
186 LLETWNEPRE-----VGSP-MEMWGGTYENRDVINRPN--LAVVNTIRAS-----G 229
215 GN-----LQAGQYPPVVLN--VPN--RLVYSAHDYATSVYPT-----WFSO- 252
230 GNDRKFLVPTNAATGLDVALNDLVIPNDSRVIVSIHAYSPYFFAMDVNGTSYWGSDY 289
253 --PTFPNNMPCGNKNGYLFNQNIAVPWLCEFGTTLQSTTDOTWLTLY-----QYLRT 306
290 DKASLTSELDLYNR-----FVKNRAVVIIEFG-----TIDKNLSSRVAHAHYAREA 339
307 AQYGADSFQWTFWSNPDSDGTGIL 332
340 VSRGIAVFWMDNGYINPGDAETVALL 365

RESULT 12
GUNI_RUMAL STANDARD; PRT; 406 AA.
ID GUNI_RUMAL
AC P16216;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EG-I).
GN EG I.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.
RX STRAIN=F-40;
RX MEDLINE=90078126; PubMed=2687251;
RA Ohmura K., Kajino T., Kato A., Shimizu S.;
RT "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene."
RL J. Bacteriol. 171:6771-6775(1989).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30528; AAA26469.1; -
CC PIR; A43722; A43722.
CC HSPG; P17901; IEDG.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00150; cellulase; 1.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 43
FT CHAIN 44 406 ENDOGLUCANASE I.
FT ACT SITE 210 210 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 330 330 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45390 MW; 2E0172437B14FEA8 CRC64;
Query Match 7.3%; Score 145.5; DB 1; Length 406;
Best Local Similarity 22.9%; Pred. No. 0.00037;
Matches 88; Conservative 49; Mismatches 152; Indels 95; Gaps 18;
15 DANNVPVRIAGIN-----WFGPCTCNVYVHGLRSR-----DYSR 47
49 ETENVPVQSTHTNDTMTVTSAKDLVAKMTNGWNLGNTMDATAQGLGSEVSMLPLKVTNKK 108
48 SMLDOIKSLGYNTIRLPYSDDILKPTMPNSINFQMNQDLQGLTSLQVMDKIVAYAGQI 107
109 YIMDLPEAGFNVLRIPIVS-----WGNHIIIDKYTSDFAWMDRVQ---EIVNYGIDN 157
108 GLRIILDRHPCSGQSALWYTSVSE-----ATWISDLOAQAQRYKGNPTVVGFD 158
158 GLYVILNTHHEE-----WYMPKPEKDGIEEIKAVWAQ---IADRFKGYDEHLIFE 206
159 LHNPHDPACWCGDPSIDWRLAERA-----GNALVSVNPENLLIFVEGVOS---YNGDS 210
207 GLNEP-----RLRGEAEWTGTSEAREIINEYKAFVETVRASGNGNDR 251
211 YWNGGNLQAGQYPP--VVLNVP--NRLVYSAHDYAT--SVYPTQWFSPTFPNN--MPGI 262
252 CLMITGYAASAYNNLSAIELPDSKLLIIVHAYLPYSPALDYKTDYDPTAIPEL 311
263 WNKWNGYLFNQNIAVPWLCEFGTTLQSTTDOTWLTLYVYLRPTAQYGADSFQWTFWSN 322
312 F-EHLNLFISKGIPIVIGFEGTWNKENTEDR-VKCEDYLAARAAKYDIPCVWMDNYA-R 368
323 PDSGDTGILKD--DWQTVDTVK 343
369 IGENGFGLMNRADLEWYFPDLIE 392

RESULT 13
GUNE_RUMAL STANDARD; PRT; 409 AA.
ID GUNE_RUMAL
AC P23661;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EGB).
GN CELB.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SY3;
RC MEDLINE=91066933; PubMed=22505649;
RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;
RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
RT genes celsa and celsB".
RL Mol. Gen. Genet. 223:217-223(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30528; AAA26469.1; -
CC PIR; A43722; A43722.
CC HSPG; P17901; IEDG.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00150; cellulase; 1.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

RP SEQUENCE FROM N.A.
RC STRAIN=AW;
RX MEDLINE=92138626; PubMed=1735723;
RA Huang J., Schell M.A.;
RT "Role of the two-component leader sequence and mature amino acid
RT sequences in extracellular export of endoglucanase EGL from
RT Pseudomonas solanacearum.";
RL J. Bacteriol. 174:1314-1323(1992).
[2]
RP SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89291722; PubMed=2738021;
RA Huang J., Sukordhaman M., Schell M.A.;
RT "Excretion of the egl gene product of Pseudomonas solanacearum.";
RL J. Bacteriol. 171:3767-3774(1989).
[3]
RP PROCESSING.
RX MEDLINE=90307678; PubMed=2195024;
RA Huang J., Schell M.A.;
RT "Evidence that extracellular export of the endoglucanase encoded by
RT egl of Pseudomonas solanacearum occurs by a two-step process
RT involving a lipoprotein intermediate.";
RL J. Biol. Chem. 265:11628-11632(1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (probable).
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M84922; AAA61980.1; -
CC PIR; A42649; A42649.
CC InterPro; IPR001547; Glyco_hydro_5.
CC InterPro; IPR000437; Prok_lipoprot_s.
CC Pfam; PF00150; cellulase; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
CC Lipoprotein; Signal; Palmitate.
CC SIGNAL 1 19
CC PROPEP 20 45
CC CHAIN 46 426
CC LIPID 20 20
CC LIPID 20 20
CC ACT_SITE 249 249
CC ACT_SITE 361 361
CC SEQUENCE 426 AA; 45578 MW; 51E13AD442CF4A8 CRC64;
Query Match 8.4%; Score 167; DB 1; Length 426;
Best Local Similarity 23.9%; Pred.No. 8e-06;
Matches 81; Conservative 49; Mismatches 129; Indels 80; Gaps 18;
QY 35 NYVHGLRSRSDYRSMQLDQIKSLGNTIRLPYSDDLKPGTWPNISFYQMNQ--DLQGLT 92
DB 141 NYIYPSADSVTY-----YKNGKNNLRLEPFRERLQP-----TLNQVFDNELS 184
QY 93 SLQVMDKIVAYAGQIGLRIILDRHPRDCSQSALWY-----TSSVSEATWISDLQALQR 147
DB 185 RLGFVNAVATGQ---TVLLDPH-----NYARYYGNVIGSSAVPNSAYADPWRRLATQ 235
QY 148 YKGNPTVGFDLNHPHDPACWGGPDSIDWRLAAREAGNAVLSVNPVLLIFVEGVQSYN 207
DB 236 FKSNPRI-V-LGLMNEPNSM-----PTEQWLSGANAEIAIRSANASNVVFPFG-NAMT 286
QY 208 GDSYWGCGNLQAGQYPVW--LNVP-NRLVYSADHY-----ATSVYPTQWFS 252

DB 287 G-AHSMNQWYTPNGTVMKGINDPGHNVFVHQYLDGSSQCSANCVSATIGAQRLQD 345
QY 253 PTFPNNMFGIWNKNWGYLFNQNIAPVWLGFGFTTQSTTDQTLWKLTVLYVRETAQYQAD 312
DB 346 FT-----TWLRSSNGY-----RGFLGFGAASNDTCNQA-VSNMLTFVKNNADV--- 387
QY 313 SFQWTFWSWNPDSGDTGGILK--DDWQTVDVTYDGYLAP 349
DB 388 ---WTGAWWAGGFWGGMYSIEPSNGVDKPMQSVLAP 423
RESULT 11
GUNE_CLOTH STANDARD; PRT; 814 AA.
AC P10477;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase E precursor (EC 3.2.1.4) (EGE) (Endo-1,4-beta-glucanase)
DE (Cellulase E).
GN CELE.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RX MEDLINE=89137992; PubMed=3066698;
RA Hall J., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "Conserved reiterated domains in Clostridium thermocellum
RT endoglucanases are not essential for catalytic activity.";
RL Gene 69:29-38(1988).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22759; AAA23224.1; -
CC PIR; JTO347; CZCLEM.
CC HSSP; P17901; LEDG.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR001547; Glyco_hydro_5.
CC InterPro; IPR001087; Lipase_GDSL.
CC Pfam; PF00150; cellulase; 1.
CC Pfam; PF00404; Dockerin_1; 2.
CC Pfam; PF00657; Lipase_GDSL; 1.
CC PROSITE; PS00448; CLOS CELLULOSE RPT; 2.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC SIGNAL 1 34
CC CHAIN 35 814
CC ACT_SITE 193 193
CC ACT_SITE 316 316
CC DOMAIN 415 474
CC REPEAT 451 438
CC REPEAT 451 474
CC SEQUENCE 814 AA; 90244 MW; C6FA24B8D1523632 CRC64;

FT DOMAIN 281 748 CATALYTIC.
 FT DISULFID 39 133 BY SIMILARITY.
 FT ACT_SITE 503 503 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 553 553 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;
 Query Match 14.7%; Score 290.5; DB 1; Length 748;
 Best Local Similarity 23.4%; Pred. No. 3.1e-15;
 Matches 111; Conservative 63; Mismatches 136; Indels 165; Gaps 23;
 QY 7 HTSGREILDANNVVRVIRAGINWFFETCN-----YVHGLRS- 43
 Db 290 NTKGNLTQDGLLPARGC--NWFLGEGHEPNDADNPSPGAPMELYAGNWWNNVNSQSG 347
 QY 44 RDYRMLDQIKSLGYNTIRLPYSDDIL-----KPGTMPNSINFYQMNQDLOGLTSLQVMDK 99
 Db 348 RTICQTWELKQGGITMLRPIAPQTLDANDPQGSFNKXN-----HQSIQSNARQALD 403
 QY 100 IVAGQIGRLIILDRH-----RPDCSGQALWY-----TSSVS 133
 Db 404 FIKLADQNDIQIFIDIHSCSNYVGRAGRLDARPPYVDANRVGYDFTRREYSCSATNPS 463
 QY 134 EAT-----WTSIDQALA--QRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAA 182
 Db 464 SVTRIHYDKQKWLUNREIAGLSAKLVSNLIGIDVNEPYD-YTWA-----EHWGMV 516
 QY 183 ERAGNAVLSVNPILLIFVEGVQSYNGDSY-----WGGNLOQAGQY 223
 Db 517 EEAQCAINEVNNFMIIIVEGI-SANANTQDGTPTSPVPFHGSDTLNPNNGENLYEAGAN 575
 QY 224 PVNLNVP-NRLVYSAHDYATSVYPTWFSDDT-----PENNMPG 261
 Db 576 PP--NIPKRLFLFPHYTGFSVFQRFQFMDPAQTECAGLEGDEAAQARIVINPTVLEQ 633
 QY 262 IWNKNWGYLFNQNTAPVWLGFEFG-----TTLQSTTDQTLKTLVQYL 303
 Db 634 GWEHFGYLRGLVG-ILIEFGGNDWPGAKSSQADRNASHITTVNDQWQQAASFY 692
 QY 304 RPTAQYAGDSQWTFWNPDSGTGILKDDWCTVTDVTKGYLAPKKSIFPDV 358
 Db 693 K---RKGINA---CYWSMNPESADTMG-----YLTTP-----WDFV 722

RESULT 9

ID GUN1_RALSO STANDARD; PRT; 424 AA.
 AC P58599;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN EGL OR RSP0162 OR R505516.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebaud P., Whalen M., Winkler P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

CC (Probable).
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL646076; CAD17313.1; ALT_INIT.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR InterPro; IPR000437; Prok_lipoprot_5.
 DR Pfam; PF00150; cellulase; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
 KW Lipoprotein; Signal; Plasmid; Complete proteome; Palmitate.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 43 BY SIMILARITY.
 FT CHAIN 44 424 ENDOGLUCANASE.
 FT LIPID 20 20 S-diacylglycerol cysteine (By
 FT similarity).
 FT LIPID 20 20 N-palmitoyl cysteine (By similarity).
 FT ACT_SITE 247 247 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 359 359 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;
 Query Match 9.4%; Score 186; DB 1; Length 424;
 Best Local Similarity 25.5%; Pred. No. 2.5e-07;
 Matches 86; Conservative 50; Mismatches 125; Indels 76; Gaps 19;
 QY 35 NYTVHGLSRDYSRMLDQIKSLGYNTIRLPYSDDILKPGTMPNSINFYQMNQ--DLQGLT 92
 Db 139 NYIYPSADSATY-----YKNGMNLVLPFRWERLQP-----TLNQALDANELS 182
 QY 93 SLQWMDKIVAVAGQIGRLIILDRHPCDSCGQALWY-----TSSVSEATWISDQALQAR 147
 Db 183 RLTFGNVAVTAAGQ--TVLLDPH-----NYARYIGNVIGSSAVPNASAYADFWRVATQ 233
 QY 148 YKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAERAGNAVLSVNPILLIFVEGVQSYN 207
 Db 234 FKGNARVI-FGLMNEPNSM-----PTEQWLSGANAALAAIRSANASNVVFPVPG-NAWT 284
 QY 208 GDSYVWGNLOQAGQYPVV--LVNENR-LVYSAHDY-----ATSVYPTWFSDD--PT 254
 Db 285 G-AHSWNQWYGTPTNGTVNKGINDPGRNLVFEVHGYLDGSDSGSGSASCVSATIGARLQD 343
 QY 255 FPNMMPGILWKNWGYLFNQNTAPVWLGFEFGTTLQSTTDQTLKTLVQYLRPTAQYAGDSF 314
 Db 344 FTN-----WLRNGY-----RGLGFEFGAASNDTCNQ-AVANMLTFVKNADV----- 385
 QY 315 QMTFWSWNPDSGDTGILK--DDWQTVTVTKDGYLAP 349
 Db 386 -WTGWAWWAGGFWGGMYSIEPSNGVDKQPMQSVLAP 421

RESULT 10

GUN2_RALSO
 ID GUN2_RALSO STANDARD; PRT; 426 AA.
 AC P17974;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN EGL.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;


```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase B precursor (EC 3.2.1.4) (EGB) (Endo-1,4-beta-glucanase)
DE (Cellulase B).
GN CELB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=86148508; PubMed=3453102;
RA "Grepinet O., Beguin P.;
RT "Sequence of the cellulase gene of Clostridium thermocellum coding
RL for endoglucanase B.";
RL Nucleic Acids Res. 14:1791-1799 (1986).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; X03592; CAA27266.1; -.
CC F1R; A23512; C2CLEM.
CC HSSP; P54583; 1ECE.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR002048; EF-hand_5.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00150; Cellulase; 1.
CC Pfam; PF00404; Dockerin_1; 2.
CC PROSITE; PS00018; EF_HAND; UNKNOWN 1.
CC PROSITE; PS00448; GLOS_CELLULOSE_RPT; 2.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC KW SIGNAL 1 27
CC FT CHAIN 28 563 ENDOGLUCANASE B.
CC FT ACT_SITE 204 204 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
CC FT DOMAIN 502 557 2 X 24 AA APPROXIMATE REPEATS.
CC FT REPEAT 502 526 1.
CC FT REPEAT 534 557 2.
CC SEQUENCE 563 AA; 63929 MW; 866FE557041DE4B CRC64;
CC
CC Query Match 26.6%; Score 528; DB 1; Length 563;
CC Best Local Similarity 33.0%; Pred No 4.5e-34;
CC Matches 130; Conservative 57; Mismatches 159; Indels 74; Gaps 13;
CC
CC QY 1 AGGY-----W-HTSGREILDANNPVRIAGINWFGFETCNVYVHGLRSRDYSML 50
CC Db 27 AEGSYADLAEPDDDLWAVEGTNIVDKYGKWIITGANWFGFCRERMLDLSYHSDIADI 86
CC QY 51 DQKSLGVNTIRLPVSDIL--KPGTTPNSINFYQMNDLQGLTSLQVMDKIVAVAGQI 107
CC Db 87 ELVADKGINVVRMPDIADLLYANWQGIYFPSTDTSYNNPALAGLSYELFNFMLENKRV 146
CC QY 108 GLRIILDRHPDCSQS---ALWTTSSVSEATWISDLOALQIAORVKGKNTPTVVGFDLHNPH 164
CC Db 147 GIKVILDVHPETDQGNHVPNTTITETIFFKAWVWAERYIKNDTIIIGFDLKNPH 206
CC 165 -----DPACWCGGCDPSIDWRLAERAGNAVLSVNPNTLLIFVEGVSY----- 206

```


28-FEB-2003 (Rel. 41, Last annotation update)
 Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
 (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
 (Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolase)
 (1,4-beta-cellulohydrolase)].
 CELB.
 Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldicellulosiruptor.
 NCBI_TaxID=44001;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=89098398; PubMed=2789547;
 SAUL D.U., WILLIAMS L.C., LOVE D.R., CHAMLEY I.W., BERGQUIST P.I.;
 RA "Nucleotide sequence of a gene from Caldocellum saccharolyticum
 RT encoding for exocellulase and endocellulase activity.";
 RL Nucleic Acids Res. 17:439-439 (1989).
 CC CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
 CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
 CC AN ENDOGLUCANASE.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
 CC F (family 10 of glycosyl hydrolases).
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial/
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

 CC EMBL: X13602; CAA31936.1; --
 DR F1R; S02711; S02711.
 DR HSP; Q06851; INBC.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR001000; Glyco_hydro_10.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00942; CBM_3; 1
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00331; Glyco_hydro_10; 1.
 DR PRINTS; PD00134; GLHYDLASE10.
 DR ProDom; PD001947; CBD_3; 1.
 DR SMART; SMC0633; Glyco_10; 1.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat;
 KW Multifunctional enzyme; Signal.
 DR FT CHAIN 1 28
 DR FT SIGNAL 29 1039
 DR FT DOMAIN 376 416
 DR FT DOMAIN 417 570
 DR FT DOMAIN 571 618
 DR FT ACT_SITE 177 177
 DR FT ACT_SITE 285 285
 DR FT ACT_SITE 792 792
 DR FT ACT_SITE 792 792
 DR SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;
 Query Match 31.3%; Score 620.5; DB 1; Length 1039;
 Best Local Similarity 35.5%; Pred No. 5.1e-41;
 Matches 13; Conservative 68; Mismatches 135; Indels 39; Gaps 12;
 QY 6 W-HTSGREILDANNVPYRIAGINWFGPTECNVYVHGLRSRDRYSMLDQIKSLGWNITRLP 64
 DB 629 WLYVSGNKIVDKGRPWLVTGINWFGYNTGTNVFDGWSCNLKDTEAIAIRGNGLLRVP 688

PRO/SER/THR-RICH (LINKER).
CELLULOSE-BINDING (BY SIMILARITY).
PROTON DONOR.
NUCLEOPHILE.

FT	DOMAIN	401	461
FT	DOMAIN	462	562
FT	ACT_SITE	203	323
FT	ACT_SITE	323	323
FT	DISULFID	75	161
FT	DISULFID	209	212
FT	STRAND	48	50
FT	TURN	51	52
FT	TURN	53	55
FT	TURN	57	58
FT	TURN	61	62
FT	STRAND	65	68
FT	STRAND	72	72
FT	TURN	73	73
FT	TURN	75	76
FT	TURN	80	84
FT	TURN	87	96
FT	TURN	97	98
FT	STRAND	101	107
FT	HELIX	108	111
FT	TURN	113	114
FT	HELIX	127	129
FT	TURN	130	131
FT	HELIX	134	147
FT	TURN	148	149
FT	STRAND	151	158
FT	STRAND	160	160
FT	TURN	161	162
FT	STRAND	163	163
FT	HELIX	175	188
FT	TURN	189	191
FT	TURN	193	194
FT	STRAND	195	199
FT	TURN	206	207
FT	STRAND	209	209
FT	TURN	215	217
FT	STRAND	218	218
FT	HELIX	219	230
FT	TURN	231	233
FT	TURN	235	236
FT	STRAND	238	242
FT	STRAND	244	244
FT	STRAND	246	247
FT	TURN	248	249
FT	STRAND	250	251
FT	TURN	254	255
FT	STRAND	257	257
FT	TURN	259	263
FT	STRAND	274	279
FT	TURN	283	285
FT	HELIX	289	292
FT	TURN	294	299
FT	HELIX	300	307
FT	TURN	308	308
FT	HELIX	309	313
FT	TURN	314	315
FT	STRAND	319	323
FT	HELIX	331	343
FT	TURN	344	344
FT	HELIX	347	350
FT	TURN	351	352
FT	STRAND	356	359
FT	TURN	367	369
FT	STRAND	373	373
FT	TURN	375	376
FT	STRAND	380	380
FT	HELIX	382	387
FT	TURN	388	388
FT	HELIX	389	391
SQ	SEQUENCE	562 AA;	60747 MW; 84E6256406A35041 CRC64;

Query Match 99.6%; Score 1974; DB 1; Length 562;
Best Local Similarity 99.7%; Pred. No. 7.7e-148;

Matches	357;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	AGGYWHTSGREILDANNVVRVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT	60						
Db	42	AGGYWHTSGREILDANNVVRVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT	101						
QY	61	IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSQVMDKIVAYAGQIGRIILDRHRPDC	120						
Db	102	IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSQVMDKIVAYAGQIGRIILDRHRPDC	161						
QY	121	SGQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	180						
Db	162	SGQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	221						
QY	181	AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNLQAGQYFVNLNPNRLVYSAHDY	240						
Db	222	AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNLQAGQYFVNLNPNRLVYSAHDY	281						
QY	241	ATSVYPTWFSDDPTFPNNMFGIWNKNWGYLFNQNIAPVWLGEFTTILQSTTDQTLKTLV	300						
Db	282	ATSVYPTWFSDDPTFPNNMFGIWNKNWGYLFNQNIAPVWLGEFTTILQSTTDQTLKTLV	341						
QY	301	QYLRPTAQYGADSPQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV	358						
Db	342	QYLRPTAQYGADSPQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV	399						

RESULT 2.

GUN_PAEPO	STANDARD;	PRT;	397 AA.
ID	GUN_PAEPO		
AC	P23548;		
DT	01-NOV-1991 (Rel. 20, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).		
OS	Paenibacillus polymyxa (Bacillus polymyxa).		
OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.		
OX	NCBI_taxID=1406;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90170677; PubMed=2307659;		
RA	Baird S.D., Johnson D.A., Seligy V.I.;		
RT	"Molecular cloning, expression, and characterization of		
RT	endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus		
RT	circulans.";		
RL	J. Bacteriol. 172:1576-1586(1990).		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.		
CC	-!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl		
CC	hydrolases).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@sb-sib.ch).		
CC	-----		
DR	EMBL; M33791; AAA22631.1; -.		
DR	PIR; A35136; A35136.		
DR	HSSP; P54583; LECE.		
DR	InterPro; IPR001547; Glyco_hydro_5.		
DR	Pfam; PF00150; cellulase; 1.		
DR	PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase.		
FT	ACT_SITE 194 194		
FT	ACT_SITE 317 317		
FT	NUCLEOPHILE (BY SIMILARITY).		
SQ	SEQUENCE 397 AA; 44357 MW; B9C2E802C04F0A2A CRC64;		

Query Match 52.3%; Score 1036; DB 1; Length 397;
Best Local Similarity 54.2%; Pred. No. 3.1e-74;
Matches 193; Conservative 43; Mismatches 108; Indels 12; Gaps 4;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.33333 Seconds
(without alignments)
2236.934 Million cell updates/sec

Title: US-09-997-504A-12
Perfect score: 1982
Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	562	1	GUN1_ACICE
2	1036	52.3	397	1	GUN_PAEPO
3	792.5	40.0	484	1	GUN_XANCP
4	620.5	31.3	1039	1	GUN_CALSA
5	596.5	30.1	566	1	GUN_CLOTM
6	541.5	27.3	747	1	GUN_CELFI
7	528	26.6	563	1	GUNB_CLOTM
8	290.5	14.7	748	1	GUNC_PSBFL
9	186	9.4	424	1	GUN1_RALSO
10	167	8.4	426	1	GUN2_RALSO
11	147.5	7.4	814	1	GUN_CLOTM
12	145.5	7.3	406	1	GUN1_RUMAL
13	134	6.8	409	1	GUNB_RUMAL
14	134	6.8	584	1	GUNB_CLOCE
15	132	6.7	825	1	GUNB_BACS4
16	130.5	6.6	900	1	GUNH_CLOTM
17	128	6.5	448	1	GUN_CLOSA
18	123	6.2	336	1	GUNB_RUMFL
19	123	6.2	440	1	GUNB_CLOCL
20	121.5	6.1	658	1	GUN3_FIBSU
21	120	6.1	517	1	GUN_CLOLO
22	119	6.0	438	1	EXG_CANAL
23	117	5.9	501	1	YBQ6_YEAS
24	115	5.8	388	1	GUN3_HUMIN
25	115	5.8	562	1	EXG2_YEAS
26	112.5	5.7	418	1	GUN2_TRIRE
27	111.5	5.6	466	1	GUN5_THSFU
28	111.5	5.6	532	1	HEXA_DICDI
29	111	5.6	764	1	YIS7_YEAS
30	109.5	5.5	475	1	GUNA_CLOCE
31	108.5	5.5	1024	1	SVIP_STAAU
32	107	5.4	312	1	GUN4_RUMAL
33	107	5.4	515	1	GUND_CLOCL

34	106.5	5.4	505	1	GUNV_ERWCA
35	106.5	5.4	547	1	GUN1_BUTFI
36	106.5	5.4	800	1	GUN_BACSI
37	105.5	5.3	364	1	GUNB_RUMAL
38	105.5	5.3	504	1	GUNV_ERWCA
39	104.5	5.3	429	1	GUNB_BUTFI
40	104.5	5.3	459	1	GUNA_STRLI
41	103.5	5.2	445	1	SPR1_YEAS
42	103	5.2	620	1	TYRO_NEUCR
43	102.5	5.2	522	1	MAL2_DROME
44	102	5.1	444	1	GUNV_ERWCA
45	99	5.0	383	1	MANA_STRLI

ALIGNMENTS

RESULT 1
GUN1_ACICE
ID GUN1_ACICE STANDARD; PRT; 562 AA.
AC P54583; 1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)
DE (Cellulase E1) (Endocellulase E1).
OS Acidothermus cellulolyticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Frankineae; Acidothermaceae; Acidothermus.
OX NCBI_TaxID=28049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43068 / 11B;
RA Laymon R.A., Himmel M.E., Thomas S.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.
RX MEDLINE=96346058; PubMed=8718854;
RA Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplus P.A.;
RT "Crystal structure of thermostable family 5 endocellulase E1 from
RT Acidothermus cellulolyticus in complex with cellotetraose.";
CC Biochemistry 35:10648-10660(1996).
CC -!- FUNCTION: Has a very high specific activity on
CC carboxymethylcellulose.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- MISCELLANEOUS: Thermostable enzyme with an optimal temperature of
CC 81 degrees Celsius.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U33212; AAA75477.1; .
PDB; 1ECE; 14-OCT-96.
DR PDB; 1CDD; 23-JUL-99.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 562
FT DOMAIN 42 400
FT CATALYTIC

This Page Blank (uspto)

Search completed: August 2, 2004, 16:37:39
Job time : 13.3333 secs

Query Match	10.8%;	Score 214;	DB 2;	Length 526;
Best Local Similarity	23.2%;	Pred. No. 5.6e-09;		
Matches	86;	Conservative 60;	Mismatches 161;	Indels 64; Gaps 14;
Qy	8	TSGREILDANNVPYRIAGINWFGFETCNYYVHGLRSRDYRSMLDQIKSLGVTIRLPY--	65	
Db	30	TKSEWIVNNKHGRVYKLCANW-PSHLXPVVAEGLSSQPMDSISKIKDMGFNCVRLLTWPL	88	
Qy	66	---SDDTLKPGTWPNFINFYQMNDLQGLTS-----LOVMKDIYVAAQOIGLRI	111	
Db	89	ELMINDIAFNWTVYKQSFERYGLDHEHQGIYHNPIYINTPLINVFQAVVYSLGRHVMV	148	
Qy	112	ILDRHP-----DCSGQ--SALWYTSVSEATWISDLQAQRYKGNPTVCGFDLHNEPHD	165	
Db	149	ILDNHKTVPFGCCSNDNDPDAFFGPKFNPLDLMLGLKMAITFNWKNVYVMSLRNELR-	207	
Qy	166	PACWGCGDPSIDWRLAAERAGNAVLSYNPNLLIIFVEGVQSYNGDSYKMGNGNLQAGAGYVP	225	
Db	208	---GYNHTSKDWYKYMKGAEAVHTSNPNVLVLTSLGN-----FDADLSFLKDRPV	255	
Qy	226	VLNVENRLVYSAHDYATSVYPTQWFSDDPT--FPNNMPEGIMWNKMGYLFNQIAFPWLGEF	283	
Db	256	NLSFKKLVLLEHWYSFTDGTGQWKSHNVNDFCSQMFSKERRITGGFVLDDQGF-FLFLSEF	314	
Qy	284	GT-----TLQSTTDQTWLKTLVQYLRPTAQYAD-SFQWTFS-----WNPDSDGTG----	329	
Db	315	GTDRGGDLEGN-----RYNMCMLAAWAEKDLDAWAVYGVYFREGKRGVVEA	364	

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreer, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, M.C.; Franco, M.C.; Frohn
J.D.; Junqueira, M.B.; Kemper, E.L.; Kitajima, J.F.; Krieger, J.E.; Krumae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0810

Query Match 16.5%; Score 328; DB 2; Length 565;
Best Local Similarity 29.6%; Pred. No. 7.4e-18;
Matches 103; Conservative 54; Mismatches 135; Indels 56; Gaps 12;
QY 12 EILDANNV-----PVRIAGINWFGFETCNVYVHGLRSRSDYRSMIDQIKSLGNTIRL 63
Db 4 ETMDAQNVSDFTHFVEPLH--GVNWRGLETAQLHGLDQRPWRVLDQMQLGINAIRL 61
QY 64 PYSDDILKGTWPNISNFQYQMDLOGLTSLQVMDKIVAYAGQIGLRIILDRHPCD-SG 122
Db 62 PLCSDDLH-GTMTPTNLDLVR-NPDLKGRFALQIDAIDEAGKGRVRLAYHVECEPTD 119
QY 123 QSALWTTSSVSATWISDLQALAQRYKGNFTVV-GFDLHNEPHDPACGCGDPSIDWRLA 181
Db 120 GNPLLSVDESEHOMSDVQFITSYRAQKQVVMGVLDAMNHRPFQSGGSDTFDNRV 179
QY 182 AERAGNAVLSVNPMLLIFVEGV-----QSYNGDSYVWGNLQAGQVPLVNPVR-LVYS 236
Db 180 VERAAAAILAMTPEWILIGVQVGLNPCLDASAPISDDNIQ--SCHCVFLRIPARLLM 237
QY 237 AHDYATSVYPTQWFSPTFPNNMPGIWKNWGLFNQNIAPVNLGFEFG---TTLQSTTDQ 293
Db 238 PHFAGTDIDTE-----AAAGAFSGKQIVLNSLDA 267
QY 294 TWLKTILVQLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDWQTVDT 341
Db 268 TDAEQLAHRIDALLAFGR--QGYFGSMWTSQMPEGMLDNDGRTPRT 313

RESULT 13
S19652
celloextrinase C - Pseudomonas fluorescens
C:Species: Pseudomonas fluorescens
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S19652; S16849
R:Ritter, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.
Biochem. J. 279, 793-799, 1991
A:Title: The celloextrinase from Pseudomonas fluorescens subsp. cellulosa consists of a
A:Reference number: S19652; MUID:92061996; PMID:1953673
A:Accession: S19652
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-748 <PER>
A:Cross-references: EMBL:X61299
R:Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.
submitted to the EMBL Data Library, August 1991
A:Description: The celloextrinase from Pseudomonas fluorescens subsp. cellulosa consists
A:Reference number: S16849
A:Accession: S16849
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262, 'K', 293-748 <PE2>
A:Cross-references: EMBL:X61299; NID:945501; PID:945502
C:Genetics:
A:Gene: celC
C:Superfamily: bacterial cellulose-binding domain homology; glycosidase CWGW domain hom
F:38-134/Domain: bacterial cellulose-binding domain homology <BCB>

F:183-220/Domain: glycosidase CWGW domain homology <GWG>
F:39-133/Disulfide bonds: #status predicted

Query Match 14.7%; Score 290.5; DB 2; Length 748;
Best Local Similarity 23.4%; Pred. No. 9.3e-15;
Matches 111; Conservative 63; Mismatches 136; Indels 165; Gaps 23;
QY 7 HTSGREILDANNVPVRIAGINWFGFETCN-----YVHGLRS- 43
Db 290 NTQGNLTKDQQLPARCG--NWFLEGRHPEPSDADNPSGAPMELYAGNWMVWVNSQSGG 347
QY 44 RDRYSMLDQIKSLGNTIRLPIYSDDL-----KGTWPNISNFQYQMDLOGLTSLQVMDK 99
Db 348 RTIQQTWTELKQGITWLRPIAPQTLDANDPGRSPNLKN-----KQIRQSNRAQLED 403
QY 100 IVAYAGQIGLRIILDRH-----RFDSCGQSALWY-----TSSVS 133
Db 404 FIKLADQNDIQIFDIHSCSNVYVWRAGRLDAPPYVDANRVGYDFTRREYSCATNPS 463
QY 134 EAT-----WISDLOALA--QRYKGNFTVVGFDLHNEPHDPACGCGDPSIDWRLA 182
Db 464 SVTRIHAIDYKQKLANLREIAGLSAKLVGNLIGIDVFNEPYD-YTWA-----EWKGMV 516
QY 183 ERAGNAVLSVNPMLLIFVEGVQSYNGDSY-----WNGNLOQAGQY 223
Db 517 EAYCAINEVNPMLLIVEGI--SANANTQGTPTDTSVPVPHGSTDLNPNWGNLYEAGAN 575
QY 224 PVLNVP-NRLVSAHDYATSVYPTQWFSPT-----FPNMPG 261
Db 576 PP--NIPKDRLLFSPHTYGPSFVQRFMDPQAOTCEAGLEGDEAAQAARCRIVINPTVLEQ 633
QY 262 IWNKNWGLFNQNIAPVNLGFEFG-----TTLQSTTDQTLKTLVQYL 303
Db 534 GWBEHGLRLGLVG--ILIEFGGNMDWPGAKSQADRNAMSHITTVNDQWQQAASYF 692
QY 304 RPTAQYAGDSFQWTFWSNPDSDGTGILKDDWQTVDTVKDGLYAPIKTSIFDPV 358
Db 693 K---RKGINA--CYWSMNPESADTMG---W-----YLTP-----WDPV 722

RESULT 14

E86265
protein F3F19.15 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86265
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86265
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <STO>
A:Cross-references: GB:AE005172; NID:94850396; PIDN:AAD31066.1; GSPDB:GN00141
C:Genetics:
A:Gene: F3F19.15
A:Map position: 1

Query Match 12.7%; Score 252.5; DB 2; Length 522;
Best Local Similarity 24.9%; Pred. No. 5.4e-12;
Matches 93; Conservative 59; Mismatches 151; Indels 71; Gaps 18;
QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRSDYRSMIDQIKSLGNTIRLPSYD 67
Db 7 TSSRWIVDNGLRVLCVCAW--PSHLOFVVAEGLSKQPDVAKKIVEMGFNCVRLTWPL 65

Db 343 YAPWFKEKDTMTQLLDYDYDTWAYTHDQGIAPILIGEWGMDGGKQKQKWTLLRDYI 402
QY 304 RPTAQYAGDSFQWTFWSPDSDGTGILKDDQWTDVTVKDGVLAP 349
Db 403 -----VQNRHHTFWCINPNSGDTGGLGNDWSTWDEAKYALLKP 442

RESULT 10
B47093
Cellulase (EC 3.2.1.4) CenD - Cellulomonas fimi
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Cellulomonas fimi
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
C:Accession: B47093
R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD),
A:Reference number: A47093; MUID:93209933; PMID:8458833
A:Accession: B47093
A:Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-747 <MEI>
A:Experimental source: ATCC 484
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIIP:128122)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:64-745/Domain: bacterial cellulose-binding domain homology <BCB>
F:645-744/Disulfide bonds: #status predicted

Query Match 27.3%; Score 541.5; DB 2; Length 747;
Best Local Similarity 34.7%; Pred. No. 2.2e-34;
Matches 131; Conservative 59; Mismatches 142; Indels 45; Gaps 15;
QY 1 AGGGYW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYN 59
Db 40 ATGDDMLHVEGNTIVDSTGKEALLSGVNVWFGFNASERVFEGLSGNTITQIQMAQRGIN 99
QY 60 TIRLPYSDDIL---KPGTM--PNSINFYQMNQDLQGLTSQVMDKIVAYAGQIGLRIILD 114
Db 100 VVRVPVSTQLLEWKAGTFLEKN-VNTY-ANPELEGKNSLIQIPEYMLTLCQYGIKVFLD 157
QY 115 RH--RPPCSQ-SALWTSYVSEATWISDQALAQRYKGNPTVVGFDLHNEPH-----D 165
Db 158 VHSAEADNSGHVYNNWKGDIITTEDVYEGWEWAATRWKDDDTIVGADIKNEPHGTQGST 217
QY 166 PACWGCDDPSIDWRLAERAGNAVLSVNPMLLI FVEGVQSYN-----GDSY--W 212
Db 218 RAKWDGTTDXDNFKHFAETASKILANPNWLVFEGVEIYPRGVPWTSTGLTDYGTW 277
QY 213 WGNLQAGQYPPVNLV-PNRLVTSADHYATSVTPQWTF-SDTFPPNMPGINKWNGYL 270
Db 278 WGNLGRVGRDHPIDILGADQDLVYSPHDYGLVFPDQKWFQDKASLTADVWGNPNLFI 337
QY 271 PNQIAAPVLFEGFTTL-QSTTDQW---LKTLYQLRPTAQYAGDSFQWTFWSPDSDG 326
Db 338 HDEIAPLLIGEWGRGLQDPRQDKWAAALDLVAER-----LSQTFVWLNPNSG 368
QY 327 DTGGLKDDQWTDVTVK 343
Db 389 DTGGLLDDWKTWDEVK 405

RESULT 11
CZCUBM
Cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase B precursor
C:Species: Clostridium thermocellum
C:Date: 31-Dec-1999 #sequence_revision 31-Dec-1989 #text_change 18-Jun-1999

C:Accession: A23512
R:Grepinat, O.; Beguin, P.
Nucleic Acids Res. 14, 1791-1799, 1986
A:Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endogluc
A:Reference number: A23512; MUID:86148508; PMID:3453102
A:Accession: A23512
A:Molecule type: DNA
A:Residues: 1-563 <GRE>
A:Cross-references: GB:X03592; NID:g40668; PIDN:CAA27266.1; PID:g40669
C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase co
C:Genetics:
A:Gene: celB
A:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans
A:Pathway: cellulose degradation
C:Superfamily: cellulase B; Clostridium cellulase repeat homology
C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-563/Product: cellulase B #status predicted <MAT>
F:502-525/Domain: Clostridium cellulase repeat homology <CCR1>
F:534-557/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 26.6%; Score 528; DB 1; Length 563;
Best Local Similarity 31.0%; Pred. No. 1.7e-33;
Matches 130; Conservative 57; Mismatches 159; Indels 74; Gaps 13;
QY 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRML 50
Db 27 AGSVADLAEPDDMLHVEGNTIVDKYKWTITGANWFGNCRERMLLDSYHSDIADI 86
QY 51 DQIKSLGYNTIRLPYSDDIL---KPGTMPNSINFYQMNQDLQGLTSQVMDKIVAYAGQI 107
Db 87 ELVADKGINVVRMPTATLLYAWSQGIYPPSTDTSYNNPALAGLSYELFNFLENFKRV 146
QY 108 GLRIILDRHRPDCSQS---ALWTSYVSEATWISDQALAQRYKGNPTVVGFDLHNEPH 164
Db 147 GIKVILDVHSPETDQGNHYPLWYNTTITTEIFKAWVWVAERYKNDTIIIGFDLKNEPH 206
QY 165 -----DPACWGCDDPSIDWRLAERAGNAVLSVNPMLLI FVEGVQSY----- 206
Db 207 TWTGTMKIKAQSAIWDSDSNHPNNWKEVAETALALEVHPNVLIFVEGVVYPRKDGIMDD 266
QY 207 -----NGDSY--WGNLQAGQYPPVNLV-PNRLVTSADHYATSVTPQWTFSDP 253
Db 267 EFTDTSPTWNTDNYGNWGNLGRGVDPINLGRYQSLVYSPHDYGPVYEQDWFKGD 326
QY 254 TFPNN-----MPGIWNKNWGYLFNQNIAPVLMGEGFTTLQSTTDQWTLKTLVQYLRP 305
Db 327 FITANDEQAKRILYQCCWRDWAYIMEGIFPLLLGEGWGTGEG--GHPLLDLNLKYLRC 384
QY 306 TAQYAGDS---FQWTFWSPNPDSDGTGILKDD-----WQVTDVTVKDGVLAPI 350
Db 385 MRDFILENKYKLHHTFWCINIDSDADTGLFTRDEGTPPPGGRDLKWN--DNKYDNYLYPV 442

RESULT 12
B82761
Extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a56)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82761
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82761
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-565 <SIM>
A:Cross-references: GB:AE003320; GB:AE003849; NID:g9105697; PIDN:AAF83620.1; GSPDB:GN0012
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.3%; Score 620.5; DB 2; Length 915;
Best Local Similarity 35.5%; Pred. No. 1.9e-40;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLP 64
DB WLVVSGNKIVDKGRPVLTGYNWFGTNTGNTVFGVWSCKLDTLAEIANRGFNLLRVP 564
QY 65 YSDIL---KPGTMN-SINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRH--RP 118
DB ISAEILNWSQGIYKPKNINYY-VNPELEKNSLEVFQVQCKEVLKIMLDIHSIKT 623
QY 119 DCSGO-SALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDP-----ACWG 170
DB 624 DAMGHYIPWYDEKFTPEDFYKACEMITRYKXNDTIIAFDLKNEPHGKFPWQDTTFAKD 683
QY 171 CGDPSIDWRLAARAGNAVLNPNLLIFVEGVQSYNGD-----ENHIHTFWCFNANS 217
DB 684 NSTDINNWKYAAETCAKRIILNPNLLIVIEGIEAYPKDDVTWTSKSSSDYSTWWG 743
QY 218 QGAGQYPPVL-NVFNRLVYSAHDYATVYPTWFSDPFPNN--MPGIWNKNGYLFNQN 274
DB 744 RGVRYKPINLGYQNKVYSPHDYGSVYQPPWFY-PGFTKESLLQDCWRPNWAYIMBEN 802
QY 275 IAPVWLGEFGTTLQSTTDOTWLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB 803 IAPLLIGEWGHLGDAGNEKWKYLRDYII-----ENHIHTFWCFNANS 856
QY 335 DWQTVDTVKDGYLAP 349
DB 857 DFTTWDEKYSFLKP 871

RESULT 8

S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Caldocellum saccharolyticum
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C;Accession: S02711
R;Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A;Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo
A;Reference number: S02711; MUID:89098398; PMID:2789517
A;Accession: S02711
A;Molecule type: DNA
A;Residues: 1-1039 <SAU>
A;Cross-references: EMBL:X13602; NID:940645; PIDN:CAA31936.1; PID:940646
C;Genetics:
A;Gene: celB
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
F;1-28/Domain: signal sequence #status predicted <SIG>
F;25-1039/Product: cellulase #status predicted <MAT>
F;72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.3%; Score 620.5; DB 2; Length 1039;
Best Local Similarity 35.5%; Pred. No. 2.2e-40;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLP 64
DB 629 WLVVSGNKIVDKGRPVLTGYNWFGTNTGNTVFGVWSCKLDTLAEIANRGFNLLRVP 688
QY 65 YSDIL---KPGTMN-SINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRH--RP 118
DB 689 ISAEILNWSQGIYKPKNINYY-VNPELEKNSLEVFQVQCKEVLKIMLDIHSIKT 747

QY 119 DCSGO-SALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDP-----ACWG 170
DB 748 DAMGHYIPWYDEKFTPEDFYKACEMITRYKXNDTIIAFDLKNEPHGKFPWQDTTFAKD 807
QY 171 CGDPSIDWRLAARAGNAVLNPNLLIFVEGVQSYNGD-----SYWWG 217
DB 808 NSTDINNWKYAAETCAKRIILNPNLLIVIEGIEAYPKDDVTWTSKSSSDYSTWWG 857
QY 218 QGAGQYPPVL-NVFNRLVYSAHDYATVYPTWFSDPFPNN--MPGIWNKNGYLFNQN 274
DB 868 RGVRYKPINLGYQNKVYSPHDYGSVYQPPWFY-PGFTKESLLQDCWRPNWAYIMBEN 926
QY 275 IAPVWLGEFGTTLQSTTDOTWLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB 927 IAPLLIGEWGHLGDAGNEKWKYLRDYII-----ENHIHTFWCFNANS 980
QY 335 DWQTVDTVKDGYLAP 349
DB 981 DFTTWDEKYSFLKP 995

RESULT 9

A40589
cellulase (EC 3.2.1.4) - Clostridium thermocellum
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Clostridium thermocellum
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jun-2000
C;Accession: A40589; S31381
R;Lenaire, M.; Beguin, P.
J. Bacteriol. 175, 3353-3360, 1993
A;Title: Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri
A;Reference number: A40589; MUID:93273701; PMID:8501039
A;Accession: A40589
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-566 <LEW>
A;Cross-references: GB:X69390; NID:940677; PIDN:CAA49187.1; PID:940678
C;Genetics:
A;Gene: celG
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A;Pathway: cellulose degradation
C;Superfamily: cellulase B; Clostridium cellulase repeat homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;503-526/Domain: Clostridium cellulase repeat homology <CCR>
F;536-559/Domain: Clostridium cellulase repeat homology <CCR>

Query Match 30.1%; Score 596.5; DB 2; Length 566;
Best Local Similarity 34.2%; Pred. No. 7.4e-39;
Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLP 64
DB 44 WLHCKGNKIYDMYGNVWLTCANWFGNCSNCFHG-AWYDVKTILTSIADRGINLRIP 102
QY 65 YSDIL---KPG---TMPNSINFYQNMQDLOG-----LTSQVMDKIVAYAGQI 107
DB 103 ISTEILYSWMLGKPNVSVTASNNPVPVNPDPYDPTDDVKNSEIPDIINGYCKEL 162
QY 108 GLRIILDRHDPDC-SQOS-ALWY---TSS---VSEATWISDLQALQARYKGNPTVVGFD 158
DB 163 GIKWMDIHSFDANNSGHNYELTKETSTCGVVTTRKMDITLVMLADKYNDDTIIAFD 222
QY 159 LHNEPHDP-----ACWCGDPSIDWRLAARAGNAVLNPNLLIFVEGVQSY- 206
DB 223 LKNEPHGKRGYTAEPVKLLAKWDNSTDENWYAAETCAKALEVNPVKLVIEGVEQY 282
QY 207 -----NGD-----SYWWG 244
DB 283 KTEKGYTYDPTDIMGATGASFPYSAMWGNLGRVKDYPIDGLPLNSQIVYSPHDYGSV 342
QY 245 YPQTFW-SDTFFPNMPCGINWKNWGYLFNQNIAFWLGEFGTTLQSTTDOTWLKTLVQYL 303

QY 228 NVPNRLVYSAHDYATSYVQPTWSDPFPNNMFGKNNKMGYLFNQNIAVPWLGEGFTLL 287
Db 268 --PRGKVVISHVGVSPYMMDDYFKSPDPNNMPLIWEHTEGYLTDNLNYTLV-IGEGNGY 325
QY 288 QSTTDTQWLKTLVQLRPTAQYAGDSFQWTFWSNPDGDTGGILKDDQWTDVTVK 343
Db 326 EG-LDKVQDAFVKMLKKIY-----NPFYWCINPESGDTGGIFLDDKWTVNWEK 375

RESULT 5
B82759
endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82759
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-592 <SIM>
A:Cross-references: GB:AB003849; NID:g9105710; PIDN:AAF83628.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
as-Netto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.B.; Laigh
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.P.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Swasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0818

Query Match 34.5%; Score 683.5; DB 2; Length 592;
Best Local Similarity 42.6%; Pred. No. 1.2e-45;
Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFTCTNYVVGHSRDRYSRMLDQIKSLGYNITRLP 64
Db 26 YSIHGKVIDDKGN-QIQKGGVFGFTTNNVHGLWLNWKEFTIQISGMLNVRLP 84
QY 65 YSDILKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHPDCSGOS 124
Db 85 FCPASLNSNTSPSID-YNRPDLQGLSLQIMDKVVKELSDRGIVYVMDHPTDCAAS 143
QY 125 ALWTSSVSSEATWISDQALAQRYKGNPTVVDPLHNEHPDPCGCGSDISDWRLAER 194
Db 144 ELWTDYSSEKQWIDDLRFVAHRYANVPVIGLVDVKNRPHGRATWGTGDPKTDWNTAVEH 203
QY 185 AGNVLVSNPNNLLIFVEGVOSYNGDS----YWMGNLQAGQYPPVNLNP-NRLVYSAHD 239
Db 204 AAAAILEAPKWLIGVEGIGENPSCSSTIGHFWGLENLEPMDCTP-LKVPADHLLIPIHV 261
QY 240 YATSVYQTFSDPTFPNNMFGKNNKMGYLFNQNIAVPWLGEGFTLL--QSTTDTQWLK 297
Db 262 YGPDVYVQYFNSPDPFNNMAAIWDKHFHFAKAGYA-MAIGFEGYKYGEGDPRDVAQN 320
QY 298 TLVQLRPTAQYQ-ADSFQWTFWSNPDGDTGGILKDDQWTV 339
Db 321 ALVDYL---ISIGVTDIF---YWSINANSDDTGGLFRDDNNHV 357

RESULT 6

E97012
probable non-processive endoglucanase family 5, secreted, CelA homolog secreted, dockerrir
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: E97012
R;Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <KUR>
A:Cross-references: GB:AB001437; PIDN:AAK78888.1; PID:g15023812; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0912

Query Match 34.0%; Score 674; DB 2; Length 482;
Best Local Similarity 37.6%; Pred. No. 5.1e-45;
Matches 140; Conservative 71; Mismatches 127; Indels 34; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFTCTNYVVGHSRDRYSRMLDQIKSLGYNITRLP 64
Db 24 YLHSDGSKLLDDYGNQVEMTGIAWFGLETPNYCFHGLWANRLDNLINIVADNGFNTLRVP 83
QY 65 YSDILKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHPD 119
Db 84 LSELVNWQRQGVYTPPSINDY-ISPELKGNSLIQLDDVIAYSKKVGVKVMMDHRIE 142
QY 120 CSQGSALWYTSVSSEATWISDQALAQRYKGNPTVVDPLHNEHPDPC 172
Db 143 SGGQTATWYTSKYTTDDYEKCQWYLADRYKNDTVAADIFNEPHGKAYRAETSAKNDT 202
QY 173 DPSIDWLAAREAGNAVLVSNPNNLLIFVEGVQSY-----NGDSY---WVGNNLQ 219
Db 203 TDENWRVEAEKGVKKLIDINPKMLIVVEGVETTPKEGTAAAGTNPDDYVYGGWGNLRG 262
QY 220 AGQYFVVL-NVFNRLVYSAHDYATSYVQPTWF-SDPTFPNNMFGKNNKMGYLFNQNIA 277
Db 263 VKDYFVLDAPYKQVYVSPHDPYGVGSDQTFWFGDFTBQSLINDIWRPSWFIQEKNIAP 322
QY 278 VWLGEFTTLQSTTDTQWLKTLVQLRPTAQYAGDSFQWTFWSNPDGDTGGILKDDQW 337
Db 323 LLIGEGMNGMDGKNEQWMTDKALI-----SDKNMHTFWCLNANSGDTGGILEYDFK 376

QY 338 TVDVTVKDGYLAP 349
Db 377 TIDTKLALVQP 388

RESULT 7
A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum se
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Caldocellum saccharolyticum
C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C:Accession: A43802
R;Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "C
A:Reference number: A43802; MUID:91136262; PMID:2126700
A:Accession: A43802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tran
A:Molecule type: DNA
A:Residues: 1-915 <SAU>
A:Cross-references: EMBL:X13602
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Streptomycetes endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

JH0158
cellulase (EC 3.2.1.4) precursor - Xanthomonas campestris pv. campestris
N:Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase
C:Species: Xanthomonas campestris pv. campestris
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: JH0158
R:Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.
Gene 89, 53-59, 1990
A:Title: Nucleotide sequence of the engXCA gene encoding the major endoglucanase of Xanthomonas campestris
A:Reference number: JH0158; MUID:90323605; PMID:2373365
A:Accession: JH0158
A:Molecule type: DNA
A:Residues: 1-493 <GOU>
A:Cross-references: GB:M32700; NID:g155397; PIDN:AA27612.1; PID:g155398
C:Comment: Endoglucanase plays a minor role in the early stages of pathogenicity of Xanthomonas campestris
C:Genetics:
A:Gene: engXCA
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A:Pathway: cellulose degradation
C:Superfamily: Xanthomonas campestris cellulase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-474/Product: cellulase #status predicted <MAT>
F:375-399/Region: proline/threonine-rich

Query Match 40.0%; Score 792.5; DB 1; Length 493;
Best Local Similarity 46.9%; Pred. No. 2.9e-54;
Matches 161; Conservative 59; Mismatches 104; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPIAGINWFGFTCNVYVHGLSRDRYSMLDQIKSLGYNTIRLP 64
Db YSINNRSQIVDGGKVVQLKGVNVFETGNGVHGLWLNWMDMLVQMGIFGNVRLP 85

QY 65 YSDIILKPGTMPNSINFYQNMQLGLTSLQVMDKIVAYAGQIGRIILDRHPCGSGS 124
Db FCPATLRSPTMPSID-YSRNADLGLTSLQIKVIAEFNARGMYVLIDHPTDCAGIS 144

QY 125 ALNVTSSVSEATWISDLQALAQRYKGNPTVVGPDLNHEDPACWCGDPSIDWRLAER 184
Db ELMYTGSTYTAQWLADLRVFNRYKVVYVGLDLKNEPHGATWGTGNAATDNWKAER 204

QY 185 AGNAVLNVNPLLIFFVEGVO-----SYNGDSYVWGNLQAGQYPVVLANP-NRLVYSAH 238
Db GSAAVLAVAPKLLIAVEGITDNPVCGTNG-GIFWGNLQPLACTP--LNIPANRLILAPH 261

QY 239 DYATSVYPTWSDPTFPNNMGIMNKNWGYLFNQNIAPVWLGEFTLL--QSTTDQTLW 296
Db VYGEDVVFQSYFNDSNFPNNMPAIWHRHFQGFAGTH--ALLLGEFGKYGEGDARDKTWQ 319

QY 297 KTLVQYLRPTAQYGADSFQWTFWNPDSGDTGGILKDDQTV 339
Db DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

RESULT 3
E71059
probable endo-1,4-beta-glucanase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: E71059
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hailkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71059
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-458 <KAW>
A:Cross-references: GB:AF000005; NID:g3236132; PIDN:BAA30271.1; PID:g3257588
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:
A:Gene: PH1171
C:Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 37.9%; Score 751; DB 2; Length 458;
Best Local Similarity 45.0%; Pred. No. 4.6e-51;
Matches 167; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPIAGINWFGFTCNVYVHGLSRDRYSMLDQIKSLGYNTIRLPYS 67
Db TSGEE-----TTHLFGVNVWFGFETPNVYVHGLWLNWMDMLLQIKSLGFNAIRLPFCT 107

QY 68 DILKPGTMPNSINFYQNMQLGLTSLQVMDKIVAYAGQIGRIILDRHPCGSGSALW 127
Db ESVPKFGQPTGID-YSKNPDLRGLDSLQIMEKLIKAGDLGIFVLLDYHRIHGTHTLEPLW 166

QY 128 YTSSVSEATWISDLQALAQRYKGNPTVVGPDLNHEDP-----PACWCGDPS 175
Db YTEDFSEEDFINTWIEVAKRFKVMNVIGADLNKNEPHSVTSPFAAYTDGTGATWGMGNTA 226

QY 176 IDWFLAERAGNAVLSVNPMLLIFFVEGVSYN-----GDSYVWGNLQAGQYFVV 226
Db TDWNLAAERIGKAILKLVAPHMLIFVEGTQFTNPKTSSYKWKYNAMWGMNMAVKDIPV- 285

QY 227 LNVP-NRLVYSADYATSVYPTWFSIPT-FPNNMFGIMNKNWGYLFNQNIAPVWLGEFG 284
Db NLPRNKLVYSPHYGPDVYNQPVFGPAKGFDPNLPDIWVHFGYVVKLELGYSVVIGFEG 344

QY 285 TTL---QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWNPDSGDTGGILKDDQTV 339
Db KYGHGDPDRVIMQNKLVDMW---IENKFCDFP---YWSNPNDSGDTGGILQDDWTTI 397

RESULT 4
E75142
endoglucanase PAB0632 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: E75142
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: E75142
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAB49854.1; PID:G545836;
A:Experimental source: strain Orsay

C:Genetics:
A:Gene: celB-like; PAB0632
C:Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 35.0%; Score 694.5; DB 2; Length 514;
Best Local Similarity 41.9%; Pred. No. 1.4e-46;
Matches 149; Conservative 57; Mismatches 121; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFTCNVYVHGLSRDRYSMLDQIKSLGYNT 60
Db YTAENGIIIFVQNVTTTTEKKKPLYLHGVNFGFELDXHVYVGLDKENWDXDLKQVRLGFNA 89

QY 61 IRLPYSDIILKPGTMPN--SINFQNMQLGLTSLQVMDKIVAYAGQIGRIILDRHPP 118
Db IRLPFCSEIAPDTPSPERIN-YELNPDKNLITSLMEKIIIEVANSIGLYLLDVHRI 148

QY 119 DCSGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGPDLNHEDPACWCGDPSIDW 178
Db GCEETPLWYNTENYSEEQYIKDWIFLAKRFKYPNVIGADIKNEPHGAGTGDGER-DF 207

QY 179 RLAAERAGNAVLSVNPMLLIFFVEGVO-----SYNGDSYV--WGNLQAGQYFVV 227
Db RLFAEKVGREILKLVAPHMLIFVEGTQYTHVNPIDELIEKKGWTFWGNLWGVKDYFVRL 267

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds

(without alignments)

2792.154 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982

Sequence: 1 AGGGYWHITSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78.*

2: PIR1.*

3: PIR2.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1038	52.4	397	A35136	cellulase (EC 3.2.
2	792.5	40.0	493	JH0158	cellulase (EC 3.2.
3	751	37.9	458	E71059	probable endo-1,4-
4	694.5	35.0	514	E75142	endoglucanase PAB0
5	683.5	34.5	592	E82759	endo-1,4-beta-gluc
6	674	34.0	482	E97012	probable non-proce
7	620.5	31.3	915	A43802	cellulase (EC 3.2.
8	620.5	31.3	1039	S02711	cellulase (EC 3.2.
9	596.5	30.1	566	A40589	cellulase (EC 3.2.
10	541.5	27.3	747	B47093	cellulase (EC 3.2.
11	528	26.6	563	CZCLEM	extracellular endo
12	328	16.5	565	B82761	celloedextrinase C
13	290.5	14.7	748	S19652	protein F3P19.15 [
14	252.5	12.7	522	E86265	hypothetical prote
15	214	10.8	526	T51476	hypothetical prote
16	204.5	10.3	488	T51502	cellulase (EC 3.2.
17	167	8.4	426	A42649	endo-1,4-beta gluc
18	154.5	7.8	722	H96986	endo-1,4-beta-gluc
19	151.5	7.6	356	G82523	endoglucanase - Th
20	149.5	7.5	329	C72216	cellulase (EC 3.2.
21	149.5	7.5	516	JE0134	cellulase (EC 3.2.
22	147.5	7.4	814	CZCLEM	endoglucanase fami
23	147	7.4	1012	B97326	cellulase (EC 3.2.
24	145.5	7.3	406	A43722	endo-1,4-beta-mann
25	140.5	7.1	669	D72278	cellulase (EC 3.2.
26	138.5	7.0	500	S22458	endoglucanase fami
27	137	6.9	370	G97001	endoglucanase B -
28	134	6.8	409	S12018	cellulase (EC 3.2.
29	134	6.8	584	J01229	cellulase (EC 3.2.

ALIGNMENTS

RESULT 1

A35136
cellulase (EC 3.2.1.4) - Bacillus polymyxa
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus polymyxa
C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 24-Sep-1998
C:Accession: A35136
R:Baier, S.D.; Johnson, D.A.; Seligy, V.L.
J. Bacteriol. 172, 1576-1586, 1990
A:Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase
A:Reference number: A35136; MUID:90170877; PMID:2307659
A:Accession: A35136
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <BAI>
A:Cross-references: GB:M33791; GB:M33840
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Xanthomonas campestris cellulase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 52.4%; Score 1038; DB 2; Length 397;

Best Local Similarity 54.2%; Pred No. 1.3e-73;

Matches 193; Conservative 44; Mismatches 107; Indels 12; Gaps 4;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLEDOIKSLGYNTRL 63

DB 37 GYHTQGNKIVDESGKEAFAFNGLNWFGLETPNYTLHGLWSRMDMLDQVKKEGYNLRL 96

QY 64 PYSDILKFGIMPNSINEYQVNDLQGLTSQVMDKIVAYAGQIGLRLDRHRPDCSQ 123

DB 97 PYSNQLFSSSRPDSID-YHKNPDLVGNLPQIMDKLIEKAGQRIQLDRHRPGSGQ 155

QY 124 SALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183

DB 156 SELWTSQYPESRWISDWKMLADRYKNNPTVIGADLHNEPHGQASWGTGNASTDWRLAAQ 215

QY 184 RAGNAVLSVNNLLTFVEGQ---SYNGDSYWGNGLOGAGQYPVVLANVRLVYSAHDY 240

DB 216 RAGNAVLSVNNLLTFVEGQVHNVGNNSQXWGNLUTGVANYPVLDVPRVYSPHDY 275

QY 241 ATSVYPTQWFSDPTPPNNMPGINKWGYLFPNQIAPVWLGEFG--TTLOSTTDTQTLKT 298

DB 276 GPGVSSQFWENDPAPFSPNLPAIDQWTCWYISKQNIAPLVGEGFGRNVLDSSPEGKQNA 335

QY 299 LVQYLRPTAQVGAQDSFQWTFWSWNPDSGTGGILKDDWQTVDTVKDGYLAPIKSSI 354

DB 336 LVHYI-----GANNLYFTYSLNPSNGDTGGLLLDDWTTWNRPKQDMLGIMKPV 385

RESULT 2

This Page Blank (uspto)

```

; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9262
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9262

Query Match      32.9%; Score 651.5; DB 15; Length 421;
Best Local Similarity 39.9%; Pred. No. 3.4e-55;
Matches 141; Conservative 60; Mismatches 123; Indels 29; Gaps 12;

Qy 5 YWHTSGREILDANNVPRIAGINWFGPETCNVYVHGLRSRDYRSMLDQIKSLGYNTIRLP 64
Db 26 YSISHGKVVDDKGN-QIQLRGVNWFGPETGDHVVVGLWARNWKEFITQLQGMGFNAIRLP 84
Qy 65 YSDDILKPGTWPNSINFEYQMNQD-LQGLTSLQVMDKIVAYAGIQLRIILDRHRPDCSGQS 124
Db 85 FCPANLNSNTSPSSID-YSRNPDLQGLSSQILDKVVKELSDRMVYLLDHRPDCSAIS 143
Qy 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRLAAER 184
Db 144 ELWHTDSVSEKQWIDDLRFVAHRVANVPVGVIGLDVKNPHGRATWGTGDFQTDWNTAVEH 203
Qy 185 AGNAVLNVNPNLLIFVEGVQSYNGDS-----YWGGNLQAGQYPPVNLVY-NELVY 235
Db 204 AAAAILEAAPKWKVIGIEGI-----GENPICSTIGHFNGENLEPMDCTP--LKVPADHLL 257
Qy 236 SAHDYATSVYPTWFSPTFPNNPFIWKNWGYLFNQNIAPVWLGEFGTTL--QSTTDQ 293
Db 258 MPHVYGPVYVQPYFNSPDPFNNMAAIDKHFHFAKAGYA-MAIGFEGKYGEGDPRDI 316
Qy 294 TWLKTLYQYLRPTAQYG-ADSFQWTFWSNPDSCDGTGILKDDWQTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNWMTTPRDDKVK 363

```

Search completed: August 2, 2004, 16:40:49
Job time : 42.3333 secs

Best Local Similarity 41.9%; Pred. No. 2.6e-59; Mismatches 149; Conservative 57; Indels 29; Gaps 10;
 7 HTSGREILDANNV-----PVRIAGINWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNT 60
 30 YTAENGIIIVQNTTGEKPKPLVHGVSWFGFKLKHVVVGLDKRKNWDLKDKVRELGENA 89
 61 IRLPSDDILKPGTWPN--SIFYQNNQDLGLTSLQWMDKIVAYAGQIGRLIILDRHRP 118
 90 IRLPFCSIRPDTRPSPERIN-YELNPKLNTSLTEIMEKIEYANSIGHYLLDYHRI 148
 119 DCSGQALWTSVSATWISDLQALQARYKGNPTVGFDLNHPDHPACWCGGDPSPIDW 178
 149 GEEIEPLWYTNSEYEQIKWIFLAKRFKYPNVIGADIKNEPHGAGWTGDER-DF 207
 179 RLAAERAGNAVLSVNPMLIFVEGVQ-----SYNGDSYV--WGNLQAGAGYPPVL 227
 208 RLFAEKVGREILKAVSHWILFVEGTQYTHVPNIDELIEKKGWTFWGENLMGVKDPVRL 267
 228 NVPNRLVYSAHDYATSVYQTFWFSDFTFNNMGWKNWGVLFNQNIAPVWLGEFGTTL 287
 268 -PRGKVYSHVGVGSPVYMDYFKSPDFNNMPIIWEHFGVLTDLNLTIV-IGEWGNY 325
 288 QSTTDQTLKTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDMQTVTVK 343
 326 EG-LDKVQDAFVKWILKKIY-----NFFWCLNPESDGTGILFDDKWTNWEK 375

RESULT 13
 US-10-369-493-17534
 ; Sequence 17534, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 17534
 ; LENGTH: 592
 ; TYPE: PRT
 ; ORGANISM: Xylella fastidiosa
 US-10-369-493-17534

Query Match 34.5%; Score 683.5; DB 15; Length 592;
 Best Local Similarity 42.6%; Pred. No. 3.9e-58;
 Mismatches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;
 5 YWHTSGREILDANNVPVRIAGINWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 64
 26 YSISHGKVIDDKGN-QIQLKGLSWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 84
 65 YSDDLKPGTWPNSINFYQNNQDLGLTSLQWMDKIVAYAGQIGRLIILDRHRPDCSGQS 124
 85 FCPASLNSNTSPSSID-YRNPDLQGLSSLIQMDKVKVLSLDRRMYVLLDHRSDCSAIS 143
 125 ALWYTSVSSEATWISDLQALQARYKGNPTVGFDLNHPDHPACWCGGDPSPIDWRLAER 184
 144 ELWYTDYSYKQWIDDLRFVAHRYANVPGVIGLDVKNPEHGRATWGTGDKPTDWTAVEH 203
 185 AGNAVLSVNPMLIFVEGVQSYNGDS-----YWMGNLQAGAGYPPVNLVNP-NRLVYSAHD 239
 204 AAAAILEAAPKWLIGVGEISCSSTIGHFNGENLEPMDCTP-LKVPADHLLIIPHV 261
 240 YATSVYPTWFSDFTFNNMGWKNWGVLFNQNIAPVWLGEFGTTL--QSTTDQTLK 297

Db 262 YGPDVYVQPFNFDFENNMAA-WDKHFGHFAKAGYA-MAIGFEGKYGEGDPRDVAMQN 320
 QY 298 TLVOYLRLPTAQY-ADSFQWTFWSNPDSDGTGILKDDMQTV 339
 Db 321 ALVDYL---ISIGVTDTF---YMSINANSDDTGLFRDDWNHV 357
 RESULT 14
 US-10-369-493-9473
 ; Sequence 9473, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 9473
 ; LENGTH: 369
 ; TYPE: PRT
 ; ORGANISM: Xylella fastidiosa
 US-10-369-493-9473
 Query Match 33.1%; Score 656.5; DB 15; Length 369;
 Best Local Similarity 40.5%; Pred. No. 9.1e-56;
 Mismatches 143; Conservative 56; Mismatches 125; Indels 29; Gaps 12;
 5 YWHTSGREILDANNVPVRIAGINWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 64
 26 YSISHGKVIDDKGN-QIQLKGLSWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 84
 65 YSDDLKPGTWPNSINFYQNNQDLGLTSLQWMDKIVAYAGQIGRLIILDRHRPDCSGQS 124
 85 FCPANLNSNTSPSSID-YRNPDLQGLSSLIQMDKVKVLSLDRRMYVLLDHRSDCSAIS 143
 125 ALWYTSVSSEATWISDLQALQARYKGNPTVGFDLNHPDHPACWCGGDPSPIDWRLAER 184
 144 ELWYTDYSYKQWIDDLRFVAHRYANVPGVIGLDVKNPEHGRATWGTGDKPTDWTAVEH 203
 185 AGNAVLSVNPMLIFVEGVQSYNGDS-----YWMGNLQAGAGYPPVNLVNP-NRLVY 235
 204 AAAAILEAAPKWLIGVGEI---GENPICSTIGHFNGENLEPMDCTP-LKVPANHLL 257
 236 SAHDYATSVYPTWFSDFTFNNMGWKNWGVLFNQNIAPVWLGEFGTTL--QSTTDQ 293
 258 MPHVVGPVYVQYFNSPDPFNNAAIWDQHFGRFAKAGYA-MAIGFEGKYGEGDPRDI 316
 294 TWLKTLCYLRPTAQY-ADSFQWTFWSNPDSDGTGILKDDMQTV--DTVK 343
 317 AMQNAFVDYL---ISIGVTDTF---YMAQNSQSDTGMVGNWNTTTRDDKVK 363

RESULT 15
 US-10-369-493-9262
 ; Sequence 9262, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 9262
 ; LENGTH: 369
 ; TYPE: PRT
 ; ORGANISM: Xylella fastidiosa
 US-10-369-493-9262
 Query Match 33.1%; Score 656.5; DB 15; Length 369;
 Best Local Similarity 40.5%; Pred. No. 9.1e-56;
 Mismatches 143; Conservative 56; Mismatches 125; Indels 29; Gaps 12;
 5 YWHTSGREILDANNVPVRIAGINWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 64
 26 YSISHGKVIDDKGN-QIQLKGLSWFGFTTCNVVHGLRSRDRYRSMLODKISLGYNTIRLP 84
 65 YSDDLKPGTWPNSINFYQNNQDLGLTSLQWMDKIVAYAGQIGRLIILDRHRPDCSGQS 124
 85 FCPANLNSNTSPSSID-YRNPDLQGLSSLIQMDKVKVLSLDRRMYVLLDHRSDCSAIS 143
 125 ALWYTSVSSEATWISDLQALQARYKGNPTVGFDLNHPDHPACWCGGDPSPIDWRLAER 184
 144 ELWYTDYSYKQWIDDLRFVAHRYANVPGVIGLDVKNPEHGRATWGTGDKPTDWTAVEH 203
 185 AGNAVLSVNPMLIFVEGVQSYNGDS-----YWMGNLQAGAGYPPVNLVNP-NRLVY 235
 204 AAAAILEAAPKWLIGVGEI---GENPICSTIGHFNGENLEPMDCTP-LKVPANHLL 257
 236 SAHDYATSVYPTWFSDFTFNNMGWKNWGVLFNQNIAPVWLGEFGTTL--QSTTDQ 293
 258 MPHVVGPVYVQYFNSPDPFNNAAIWDQHFGRFAKAGYA-MAIGFEGKYGEGDPRDI 316
 294 TWLKTLCYLRPTAQY-ADSFQWTFWSNPDSDGTGILKDDMQTV--DTVK 343
 317 AMQNAFVDYL---ISIGVTDTF---YMAQNSQSDTGMVGNWNTTTRDDKVK 363

```

Db 167 YTEDFSEEDFINTWIEVAKRGKYNNVIGADLKNEPHSVTSPAAVTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKMGYNWGMGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTQWFSDEPT-PPNNMPCGINKWGYLFPNQNIAPVWLGEFG 284
Db 286 -NLPRNKLIVSPHYGPDVYNQYFPGAKGFPDNLPIWYHHFGYVKLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWTKLVQYLRPTAGYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 10
US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 37.9%; Score 751; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.1e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSREILDANNVPVRIAGINWFGTETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLPYSD 67
Db 54 TSGBE-----TPHIFGVNWFGETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQNMQLQGLTSLOVMDKIVAYAGQICLRILDRHRPDCSGQSALW 127
Db 108 ESVKFGTQPID-YSKNPDLRGLDSLQIMEKIKKAGDGLIFVLLDYHRICTHIEPLW 166
Qy 128 YTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRGKYNNVIGADLKNEPHSVTSPAAVTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKMGYNWGMGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTQWFSDEPT-PPNNMPCGINKWGYLFPNQNIAPVWLGEFG 284
Db 286 -NLPRNKLIVSPHYGPDVYNQYFPGAKGFPDNLPIWYHHFGYVKLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWTKLVQYLRPTAGYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 11
US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 37.9%; Score 751; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.1e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSREILDANNVPVRIAGINWFGTETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLPYSD 67
Db 54 TSGBE-----TPHIFGVNWFGETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQNMQLQGLTSLOVMDKIVAYAGQICLRILDRHRPDCSGQSALW 127
Db 108 ESVKFGTQPID-YSKNPDLRGLDSLQIMEKIKKAGDGLIFVLLDYHRICTHIEPLW 166
Qy 128 YTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRGKYNNVIGADLKNEPHSVTSPAAVTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKMGYNWGMGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTQWFSDEPT-PPNNMPCGINKWGYLFPNQNIAPVWLGEFG 284
Db 286 -NLPRNKLIVSPHYGPDVYNQYFPGAKGFPDNLPIWYHHFGYVKLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWTKLVQYLRPTAGYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

```

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20347
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20347

Query Match 37.9%; Score 751; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.1e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSREILDANNVPVRIAGINWFGTETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLPYSD 67
Db 54 TSGBE-----TPHIFGVNWFGETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQNMQLQGLTSLOVMDKIVAYAGQICLRILDRHRPDCSGQSALW 127
Db 108 ESVKFGTQPID-YSKNPDLRGLDSLQIMEKIKKAGDGLIFVLLDYHRICTHIEPLW 166
Qy 128 YTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRGKYNNVIGADLKNEPHSVTSPAAVTDGTGATWGMGNA 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKMGYNWGMGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTQWFSDEPT-PPNNMPCGINKWGYLFPNQNIAPVWLGEFG 284
Db 286 -NLPRNKLIVSPHYGPDVYNQYFPGAKGFPDNLPIWYHHFGYVKLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWTKLVQYLRPTAGYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGCGDPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 12
US-10-369-493-21618
; Sequence 21618, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21618
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21618

Query Match 35.0%; Score 694.5; DB 15; Length 514;

```



```
QY 181 AAERAGNAVLSVNPMLLIIFVEGVQSYNGDSYWGNGNLOAGQYPPVVLNPNRLVYSADHY 240
DB 222 AAERAGNAVLSVNPMLLIIFVEGVQSYNGDSYWGNGNLOAGQYPPVVLNPNRLVYSADHY 281
QY 241 ATSVYPTWFSPTFPNNPGLWKNWGYLFNONTAPVWLGEFGTTLQSTDDQWLKTLV 300
DB 282 ATSVYPTWFSPTFPNNPGLWKNWGYLFNONTAPVWLGEFGTTLQSTDDQWLKTLV 341
QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
DB 342 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 399

RESULT 2
US-10-360-101-212
; Sequence 212, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouwers, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 212
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of cellulase
US-10-360-101-212

Query Match 52.3%; Score 1036; DB 15; Length 397;
Best Local Similarity 54.2%; Pred. No. 4.3e-93;
Matches 193; Conservative 43; Mismatches 108; Indels 12; Gaps 4;

QY 4 GYHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLSRDYSMLDQIKSLGVTNIRLP 63
DB 37 GYHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLSRDYSMLDQIKSLGVTNIRLP 96
QY 64 YSDDILKPGTWPNSINFQNMNDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQ 123
DB 97 PYSNQLFDSSRRPDSID-YHKNFDLVGLNPIQIMDKLIEKAGQGIQIILDRHRPDCSGQ 155
QY 124 SALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWELAAE 183
DB 156 SELWYTSQYPSERWISDWRMLADRYKKNPTVIGADLHNEPHGQASWGTGNASTDWELAAQ 215
QY 184 RAGNAVLSVNPMLLIIFVEGVQ---SYNGDSYWGNGNLOAGQYPPVVLNPNRLVYSADHY 240
DB 216 RAGNAVLSVNPMLLIIFVEGVQDHVNVQGNNSQYWGNGNLTGVANYPVLDVNPVVSYPHY 275
QY 241 ATSVYPTWFSPTFPNNPGLWKNWGYLFNONTAPVWLGEFGTTLQSTDDQWLKTLV 298
DB 276 GPGVSSQWPNDFAPFNSLPAIWDQWGYISKONIAFVLVGEFGGNGVNDLSCPEGKQWNA 335
QY 299 LVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIF 354
DB 336 LVHYI-----GANNLYFTVSLNPNSPSDGTGGILLDDWTTWRRPKQDMLGRMKPV 395

RESULT 3
US-10-369-493-16379
; Sequence 16379, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16123
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16379

Query Match 39.6%; Score 785.5; DB 15; Length 483;
Best Local Similarity 46.6%; Pred. No. 2.6e-68;
Matches 160; Conservative 59; Mismatches 105; Indels 19; Gaps 9;

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldmann, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16379
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16379

Query Match 40.0%; Score 792.5; DB 15; Length 493;
Best Local Similarity 46.9%; Pred. No. 5.6e-69;
Matches 161; Conservative 59; Mismatches 104; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLSRDYSMLDQIKSLGVTNIRLP 64
DB 26 YSINNRSQIVDDSGKVYQLKGVNVFGETGNVHVMHGLWARNKMDMIVQMGLGFNAVRLP 85
QY 65 YSDDILKPGTWPNSINFQNMNDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQ 124
DB 86 FCPATLRSDTMEASID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHHHTPDCAGIS 144
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWELAAE 184
DB 145 ELWYTSYSTEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWELAAE 204
QY 185 AGNAVLSVNPMLLIIFVEGVQ---SYNGDSYWGNGNLOAGQYPPVVLNPNRLVYSADHY 238
DB 205 GSAVLAVALAPKWLIAVEGITENPVCSING-GIFWGNLQPLACTP--LNIPANRLLLAPH 261
QY 239 DYATSVYPTWFSPTFPNNPGLWKNWGYLFNONTAPVWLGEFGTTLQSTDDQWLKTLV 296
DB 262 VYGPVDFVQSYFNDSPFNPNPFAIWERHFGQFAGTH--ALLLGFQKYGEGDARDKTIWQ 319
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTV 339
DB 320 DALVKYLR---SKGIN--QGFYWSNPNSPSDGTGGILLRDDWTSV 357

RESULT 4
US-10-369-493-16123
; Sequence 16123, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldmann, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16123
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16123

Query Match 39.6%; Score 785.5; DB 15; Length 483;
Best Local Similarity 46.6%; Pred. No. 2.6e-68;
Matches 160; Conservative 59; Mismatches 105; Indels 19; Gaps 9;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds
(without alignments)

2716.899 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982

Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTVKDGLYAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 segs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	562	9	US-09-981-900B-5
2	1036	52.3	397	15	US-10-360-101-212
3	792.5	40.0	493	15	US-10-369-493-16379
4	785.5	39.6	483	15	US-10-369-493-16123
5	785.5	39.6	501	15	US-10-369-493-15740
6	785.5	39.6	518	15	US-10-369-493-15373
7	785.5	39.6	535	10	US-09-927-827-50
8	758	38.2	553	9	US-09-888-224-2
9	751	37.9	458	14	US-10-293-344A-2
10	751	37.9	458	15	US-10-369-493-1281
11	751	37.9	458	15	US-10-369-493-20347
12	694.5	35.0	514	15	US-10-369-493-21618
13	683.5	34.5	592	15	US-10-369-493-17534
14	656.5	33.1	369	15	US-10-369-493-9473
15	651.5	32.9	421	15	US-10-369-493-9262

16	574	29.0	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	455.5	23.0	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	281.5	14.2	353	15	US-10-369-493-15733	Sequence 15733, A
19	281.5	14.2	353	15	US-10-369-493-16119	Sequence 16119, A
20	281.5	14.2	582	10	US-09-927-827-48	Sequence 48, Appl
21	249.5	12.6	533	12	US-10-424-599-176232	Sequence 176232, A
22	246	12.4	375	15	US-10-369-493-12597	Sequence 12597, A
23	237	12.0	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	195.5	9.9	438	12	US-10-425-114-66753	Sequence 66753, A
25	194.5	9.8	314	15	US-10-369-493-15726	Sequence 15726, A
26	194.5	9.8	314	15	US-10-369-493-16112	Sequence 16112, A
27	194.5	9.8	337	15	US-10-369-493-15359	Sequence 15359, A
28	186	9.4	644	16	US-10-437-963-139883	Sequence 139883, A
29	182.5	9.2	917	16	US-10-437-963-181484	Sequence 181484, A
30	181.5	9.2	375	10	US-09-917-378-3	Sequence 3, Appl
31	181.5	9.2	375	10	US-09-917-378-6	Sequence 6, Appl
32	181.5	9.2	762	10	US-09-917-378-1	Sequence 1, Appl
33	179	9.0	356	10	US-09-917-378-7	Sequence 7, Appl
34	176.5	8.9	346	10	US-09-917-378-8	Sequence 8, Appl
35	176	8.9	517	12	US-10-107-431-259	Sequence 439, App
36	168.5	8.5	558	16	US-10-437-963-185049	Sequence 185049, A
37	157.5	7.9	533	12	US-10-424-599-150988	Sequence 150988, A
38	154.5	7.8	722	12	US-10-282-122A-51993	Sequence 51993, A
39	152.5	7.7	268	12	US-10-424-599-284543	Sequence 284543, A
40	151.5	7.6	356	15	US-10-369-493-17756	Sequence 17756, A
41	150	7.6	217	16	US-10-437-963-112718	Sequence 112718, A
42	149.5	7.5	503	12	US-10-424-599-150986	Sequence 150986, A
43	146.5	7.4	315	15	US-10-369-493-9177	Sequence 9177, Ap
44	146.5	7.4	518	10	US-09-769-734-56	Sequence 56, Appl
45	144	7.3	362	15	US-10-369-493-12370	Sequence 12370, A

ALIGNMENTS

RESULT 1
US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Magbool, Shahina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE TO FERMENTABLE SUGARS
; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981,900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match 99.6%; Score 1974; DB 9; Length 562;
Best Local Similarity 99.7%; Pred No. 3, 1e-185;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLRSRDYRMLDQIKSLGNT 60
Db 42 AGGGYWHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLRSRDYRMLDQIKSLGNT 101
QY 61 IRLPYSDDLKPGTWPNSINFYQNMQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
Db 102 IRLPYSDDLKPGTWPNSINFYQNMQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 161
QY 121 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWGCGDPSIDWRL 180
Db 162 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWGCGDPSIDWRL 221

Search completed: August 2, 2004, 16:38:34
Job time : 15.6667 secs

```
; ORGANISM: Paenibacillus pabuli
US-09-797-464A-7

Query Match      6.9%; Score 137; DB 4; Length 348;
Best Local Similarity 22.6%; Pred. No. 6.5e-05;
Matches 67; Conservative 48; Mismatches 96; Indels 86; Gaps 16;

QY 49 MLDQIKSLGYNTIRLP--YSDDILKPGTMDN-SINFYQMNQDLQGLTSLQWMDKIVAYAG 105
Db 34 LIKKVKAAGFKSIRIPVSLNNI---GSAPTYTINAAMLNR-----IQQVVDYAY 80

QY 106 QIGLRITLDRHPCDQSGSALWYTSVSEATWISD-----LQALAORYKGNPTVVGFDLH 160
Db 81 NEGLYVLIINHGDYNSVQGGWLLVNGGNGTAKKKYKVKWQOIAFKFSYNDHLIFESM 140

QY 161 NEHPDPACWCGDPSIDWRLAERAGNAVLNPNLL--IFVEGVQSY--NGDSYW---- 212
Db 141 NEVFED---GNVGNP-----NSAYTNLNAYNQIFVDTVRQTGGNNNAWLLVP 185

QY 213 -WGNLQ-GAGQYPPVVLNVN-----RLVYSAHDY-----AT 242
Db 186 GWNTNIDFTVGNYPALPTDNFRSSAIPSSQKRIMISAHYSPWDFAGEENGNIQTWGAT 245

QY 243 SVYP---QTFSDPTFPNNMFGIWNK--NWGYLFNQNIAPVNLGBEGTTLQSTTDDT 294
Db 246 ATNPAKSTWQEDYLESOFKSMYDKFTQGY-----PWWIGFSGIDKTSYDST 295
```

```
RESULT 15
US-09-797-464A-2
; Sequence 2, Application US/09797464A
; Patent No. 6630340
; GENERAL INFORMATION:
; APPLICANT: Wilting, Reinhard
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Schuelein, Martin
; TITLE OF INVENTION: Family 5 Xyloglucanases
; FILE REFERENCE: 6073.200-US
; CURRENT APPLICATION NUMBER: US/09/797,464A
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Paenibacillus pabuli
US-09-797-464A-2
```

```
Query Match      6.9%; Score 136; DB 4; Length 395;
Best Local Similarity 22.6%; Pred. No. 9.8e-05;
Matches 67; Conservative 48; Mismatches 96; Indels 86; Gaps 16;

QY 49 MLDQIKSLGYNTIRLP--YSDDILKPGTMDN-SINFYQMNQDLQGLTSLQWMDKIVAYAG 105
Db 74 LIKKVKAAGFKSIRIPVSLNNI---GSAPTYTINAAMLNR-----IQQVVDYAY 120

QY 106 QIGLRITLDRHPCDQSGSALWYTSVSEATWISD-----LQALAORYKGNPTVVGFDLH 160
Db 121 NEGLYVLIINHGDYNSVQGGWLLVNGGNGTAKKKYKVKWQOIAFKFSYNDHLIFESM 180

QY 161 NEHPDPACWCGDPSIDWRLAERAGNAVLNPNLL--IFVEGVQSY--NGDSYW---- 212
Db 181 NEVFED---GNVGNP-----NSAYTNLNAYNQIFVDTVRQTGGNNNAWLLVP 225

QY 213 -WGNLQ-GAGQYPPVVLNVN-----RLVYSAHDY-----AT 242
Db 226 GWNTNIDFTVGNYPALPTDNFRSSAIPSSQKRIMISAHYSPWDFAGEENGNIQTWGAT 285

QY 243 SVYP---QTFSDPTFPNNMFGIWNK--NWGYLFNQNIAPVNLGBEGTTLQSTTDDT 294
Db 286 STNPAKSTWQEDYLESOFKSMYDKFTQGY-----PWWIGFSGIDKTSYDSS 335
```

US-08-525-697-2
 ; Sequence 2, Application US/08525697
 ; Patent No. 5795764
 ; GENERAL INFORMATION:
 ; APPLICANT: Christgau, Stephan
 ; APPLICANT: Andersen, Lene N
 ; APPLICANT: Kauppinen, Sakari
 ; APPLICANT: Helld-Hansen, Hans P
 ; APPLICANT: Dalboege, Henrik
 ; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/525,697
 ; FILING DATE: 21-SEP-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Harrington, James J.
 ; REGISTRATION NUMBER: 38,711
 ; REFERENCE/DOCKET NUMBER: 4004.204-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: DK 0486/93
 ; FILING DATE: 30-APR-1993
 ; CLASSIFICATION: 435
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 377 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-525-697-2

Query Match 9.0%; Score 178.5; DB 1; Length 377;
 Best Local Similarity 23.4%; Pred. No. 7.4e-09;
 Matches 83; Conservative 56; Mismatches 115; Indels 101; Gaps 21;

Qy 24 AGIN-WFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIRLPYSDDLKPGTWPNSINF 81
 Db 50 AGTNSYWGIFLTNNDDVLMVLSQLAASDLKILRWGFDVNTKPTD-----GTV-----W 99

Qy 82 YQMNQD-----LQGLTSLQWMDKIYVAGQIGRLILD--RHRPDCSQSAL----- 126
 Db 100 YQLHANGTSTINTGADGLRDLVYVTAQKYGKLIINFNEWTDYCGMOAYTAAGAA 159

Qy 127 ---WYTSVSEATWISDLQALQRYKGNPTVWGLDHLNEPHDPACMGCGDPSIDWRLAAE 183
 Db 160 QTDFTYNTALQAAKNYKAVSRYSAAIAFAWELANEPR---CQGC-DTSVLYNNWISD 215

Qy 184 RAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNNLQAGQYFVV-----LNVNPNLVVSAHD 239
 Db 216 -TSKYIKLSKSHLVTI-----GDE-FGFLVDSDGSYPYVYVGEGLNFTKMLGISTID 266

Qy 240 YAT-SVYPTQWFSDFPTFNPNMGIWKNWGLFNQNLIA-----PVMLEGFETT----- 286
 Db 267 FGLHLVYDSDSGTSDYD-----WGNW---ITAHAAACKAVGKPCLLIEYGVTSNHCA 315

Qy 287 LQSTTDQ-----WLKTLVQLRPTAQAGDSFQWTF-WSWNPDSGDT 328
 Db 316 VESPWQQTAGNATGISGLYW-----QYGT-----TFSMQSPNDGNT 353

RESULT 13
 US-09-797-464A-4
 ; Sequence 4, Application US/09797464A
 ; Patent No. 6630340
 ; GENERAL INFORMATION:
 ; APPLICANT: Wiltting, Reinhard
 ; APPLICANT: Bjornvad, Mads Eskelund
 ; APPLICANT: Kauppinen, Markus Sakari
 ; APPLICANT: Schulein, Martin
 ; TITLE OF INVENTION: Family 5 xyloglucanases
 ; FILE REFERENCE: 6073.200-US
 ; CURRENT APPLICATION NUMBER: US/09/797,464A
 ; CURRENT FILING DATE: 2002-02-19
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 400
 ; TYPE: PRT
 ; ORGANISM: Paenibacillus sp.
 ; US-09-797-464A-4

Query Match 7.1%; Score 141; DB 4; Length 400;
 Best Local Similarity 22.5%; Pred. No. 3.3e-05;
 Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;

Qy 3 GGYWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIR 62
 Db 48 GAGWNLGNQLEATVNGVSETA---W-----GNPVV-----TPELIKKVKAAAGFKTIR 92

Qy 63 LP--YSDDLKPGTMPN-SINFYQMNQDLOGTSLQWMDKIYVAGQIGRLILDHRPDRPD 119
 Db 93 IPVSYLNHI---GSAPNVTINAANLNR-----VQTVVDYAYNEGVLVWINIHGDG 139

Qy 120 CSQSALWYTSVSEATWISD-----LQALAORYKGNPTVWGLDHLNEPHDPACMGCGDP 174
 Db 140 YNSIFGGWLLVNGSNQAAIKYKQVQOQIATKPSNNYNERLIFESMNEVFD---GNYGNP 196

Qy 175 SIDWRLAAERAGNAVLNVNPNLL--IFVEGVQSY--NGDSYW-----WGNLQ-CAGQYP 224
 Db 197 -----NAAAYANLNAVYQIFVDVTQGTGNNNARWLLIPGWNNTIDYTVGNYG 244

Qy 225 VVL-----NVPN---RLVYSAHDY-----ATSVYV---QTFWSDP 253
 Db 245 FALPTDHFPSAIPSSQKRIMISAHYSPWDFAGEENGNIQWGAATNPSKSTWQGD 304

Qy 254 TFPNNMPTGWNK--NWGYLFNQNIAPVWLGFBGT---TLQSTTDQTLWLTQLVQLRPTA- 307
 Db 305 YLNAQFKSYDKFVTOGY-----PVVIGBFGSIDKATYSTNNVYRQAYAKAVTATAK 357

Qy 308 QYGADSPQWTFWSMNPDSGDTG 329
 Db 358 KYGAVPVYV-----DNGHNG 372

RESULT 14
 US-09-797-464A-7
 ; Sequence 7, Application US/09797464A
 ; Patent No. 6630340
 ; GENERAL INFORMATION:
 ; APPLICANT: Wiltting, Reinhard
 ; APPLICANT: Bjornvad, Mads Eskelund
 ; APPLICANT: Kauppinen, Markus Sakari
 ; APPLICANT: Schulein, Martin
 ; TITLE OF INVENTION: Family 5 xyloglucanases
 ; FILE REFERENCE: 6073.200-US
 ; CURRENT APPLICATION NUMBER: US/09/797,464A
 ; CURRENT FILING DATE: 2002-02-19
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 7
 ; LENGTH: 348
 ; TYPE: PRT

QY 338 TVDTVKDGYLAP 349
Db 562 TWDEQKYNFLKP 573

RESULT 10

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 1426 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e

SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 31.1%; Score 617; DB 3; Length 1426;

Best Local Similarity 35.5%; Pred. No. 3e-50;

Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNVPVRIAGINWFFETCNVYVHGLSRDYSRMLDQIKSLGVNTIRLPYSDDI 69

Db 1022 GNKIVDXGKPFVMTGVNWFNGTGTNVFDGVNSCNLSALAIANRGNLLRVPISAEI 1081

QY 70 L---KPTMEN-SINFQNMQDLQGLTSLOVMKIVAYAGQIGLRILDRH--RPCSQ 123

Db 1082 ILNWSKGIYKPKPINIY-VNPELEGLTSLEVFVVKTKVEGLKIMLDIHSKTDAMGH 1140

QY 124 -SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDP-----ACWGCDFPS 175

Db 1141 IYPWYTDITPEDIYKACEMITERKNDITIVAFDLKNEPHGKPMQDSVFAKWDNSTDI 1200

QY 176 IDWRLAARAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWNGNLCQAGQ 222
Db 1201 NNMKYAAETCAKRLAKPNMLIVIGIEAYPKDDVTWTSKSSSDYVSTWGGNLRGVKK 1260
QY 223 YPVVL-NVFNRLVYSAHDYATSYPTQW-----PSDPTFPNNMFGIWNKNWGYLFNQNIAP 277
Db 1261 YPINLGOYQNKVYSPHDYGLVYQPFYQPFYGTGKDTLYND---CWRDNWYIINDNGIAP 1317
QY 278 VMLGEFGTILQSTTDQTLKTLVQLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQ 337
Db 1318 LLIGEWGGLDGGDNKXWYILADYII-----ENHIHTFWCYNANSNGDTGLGLGVDFPS 1371
QY 338 TVDTVKDGYLAP 349
Db 1372 TWDEQKYNFLKP 1383

RESULT 11

US-08-276-213-2

; Sequence 2, Application US/08276213

; Patent No. 5536655

; GENERAL INFORMATION:

; APPLICANT: Thomas, Steven

; APPLICANT: Layton, Robert

; APPLICANT: Himmel, Michael

; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: USA

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/276,213

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: NREL IR# 94-08

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303)231-1000

; TELEFAX: (303)231-1098

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 38 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; US-08-276-213-2

Query Match 10.3%; Score 205; DB 1; Length 38;

Best Local Similarity 97.4%; Pred. No. 6.6e-13;

Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYWHHTSGREILDANNVPVRIAGINWFFETCNVYV 38

Db 1 AGGGYWHHTSGREILDANNVPVRIAGINWFFETCNVYV 38

RESULT 12

199 TDNLAAERIGKALIKVAPHWLIIEVEGTQFTNPKTSSYKMGYNAMWGNLMAVKDYPV- 257
227 LNVP-NRLVYSAHDYATSVYPTQWFSDBT-FPNMNGIWNKNWGYLFNQIAPVWLGEFG 284
258 -NLPNKLVSPHYVGPVQNPYFGAKGFPDNLFDIWHHGYVVKLEGLGYSVWIGFEG 316
285 TTL---QSTTDQWTKLVQYLRTAQYAGDSFQWTFWSPNPSGDTGGILKDDQWTV 339
317 GKYGHGDPDRVWQNKLVDMW--IENKFCDF--YWSNPNPSGDTGGILQDDWTI 369

RESULT 8

US-09-869-197-5

; Sequence 5, Application US/09869197

; Patent No. 6566113

; GENERAL INFORMATION:

; APPLICANT: TAKAYAMA, Masanori

; APPLICANT: UNEDA, Kahoko

; APPLICANT: KOYAMA, No. 6566113uto

; APPLICANT: ASADA, Kiyoza

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: POLYPEPTIDES

; FILE REFERENCE: TAKAYAMA-6

; CURRENT APPLICATION NUMBER: US/09/869,197

; PRIOR FILING DATE: 2001-06-25

; PRIOR APPLICATION NUMBER: PCT/JF99/07009

; PRIOR FILING DATE: 1999-12-14

; PRIOR APPLICATION NUMBER: JP 366237/1998

; PRIOR FILING DATE: 1998-12-24

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 5

; LENGTH: 458

; TYPE: PRT

; ORGANISM: Pyrococcus horikoshii OT3

US-09-869-197-5

Query Match 37.9%; Score 751; DB 4; Length 458;

Best Local Similarity 45.0%; Pred. No. 6.8e-64;

Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

8 TSREILDANNVPVRIAGINWFGFETCNVYVHGLRSRSDYRSMLEQIKSLGYNIRLPYS 67

54 TSRES-----TPHFGVNWFGFETPNVHVGWGLKSNWEDMLLQIKSLGFNAIRLPFCT 107

68 DILKPGWPNISINFYQMNQDGLTSLQVMDKIVAYAGQIGRLIILDRHPRDCSGQSAW 127

108 ESVKPGTQPIGID-YSRNPDRLGDLDSQIWEKIIKKAGDLGIFVLLDHYHRRIGCTHIEPLW 166

128 YTSVSVATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175

167 YTEPESBEDPNTWIEVAKRGKWNVIADLKNEPHSVTSPAAAYTDGTGATWGNPA 226

176 IDWRLAERAGNAVLISVNPILLIFVEGVQSYN-----GSIYWGGLNQGAGQYPVV 226

227 TDNLAAERIGKALIKVAPHWLIIEVEGTQFTNPKTSSYKMGYNAMWGNLMAVKDYPV- 285

227 LNVP-NRLVYSAHDYATSVYPTQWFSDBT-FPNMNGIWNKNWGYLFNQIAPVWLGEFG 284

286 -NLPNKLVSPHYVGPVQNPYFGAKGFPDNLFDIWHHGYVVKLEGLGYSVWIGFEG 344

285 TTL---QSTTDQWTKLVQYLRTAQYAGDSFQWTFWSPNPSGDTGGILKDDQWTV 339

345 GKYGHGDPDRVWQNKLVDMW--IENKFCDF--YWSNPNPSGDTGGILQDDWTI 397

RESULT 9

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997TUS001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 616 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 31.1%; Score 617; DB 3; Length 616;

Best Local Similarity 35.5%; Pred. No. 8.6e-51;

Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

10 GREILDANNVPVRIAGINWFGFETCNVYVHGLRSRSDYRSMLEQIKSLGYNIRLPYSDDI 69

212 GNKIVDKGKPVLAGVNVNFGTCTNVFDGVWSCNLKSALAEIANRGFNLLRVPISAEI 271

70 L---KPGIMPN-SINFYQMNQDGLTSLQVMDKIVAYAGQIGRLIILDRH--RDCSGQ 123

272 ILNWKGIYKPNINNY-VNPELEGLTSLEVFDFVVKTCKEVGLKIMLDIHSAKTDAMGH 330

124 -SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----ACWCGDPS 175

331 IYVWYTDITPEDYKACEMITERKYKNDTIVAFDLKNEPHGKWPQDSVFAKWDNSTDI 390

176 IDWRLAERAGNAVLISVNPILLIFVEGVQSYN-----SYWNGNLQAGQ 222

391 NNWYAAETCAKRIIAKPNMLIVIEGIAVPKDDVTWTSKSSSYSTWNGNLRGVK 450

223 YPVWL-NVENRLVYSAHDYATSVYPTQW---FSDPTFPNNMPPGIWNKNGWLFNQNIAP 277

451 YPINLQYQNKVYSPHYDYGPLVYQPPWFYPGFTKDTLYND---CWRDNWYIMDNGIAP 507

278 VWLGEFTTLOSTTQOTWTKLVQYLRTAQYAGDSFQWTFWSPNPSGDTGGILKDDQW 337

508 LLIGWGGYLDGGDNKRWNTYLRDYII-----ENHHHTFCWYNANSNGDTGGLVGYDFS 561

Query Match 38.2%; Score 758; DB 3; Length 553;
Best Local Similarity 45.2%; Pred. No. 1.9e-64;
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNPVRIAGINWFGFETCNVYVHGLRSRYSRMLDQIKSLGYNTRLRPSYDDILKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPCTQSVKP 63

QY 73 GTPNSINFYQMNQDLOGLSLQVMDKIVAYAGOIGRLIILDRHRPDCSGQSALWYTSV 132
DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEXIIKKAGDLGIFVLLDYHRIICNFIEPLWYTSF 122

QY 133 SEATWISDLQALQRYKGNPTVVGFDLHNEPDP-----ACWCGDPSIDWRL 180
DB 123 SEQDYINTWVEAQRFGKYNWIGADLKNPHESSPAPAAAYTDGSGATWGMGNATDNL 182

QY 181 AAERAGNAVLSVNPMLLIIFVEGVQ-----SYN-GDSYWGNGNLQAGQYPPVVLNVP- 230
DB 183 AAERIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWHNAWGNLGMVRYKYPV--NLPR 240

QY 231 NRVLYSAHDYATSVYPTWFSDP--TFNNMFGIWNKNGYLFNQNIAPVWLGEFGTTL- 287
DB 241 DKVYSPQVYSGVYDQPYF-DPGEFPDNLPEIWHFHYKVLDLGYFVWIGFEGKYG 299

QY 288 --QSTTDQTLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSWNPNSGDTGGILKDDWTTI 348

RESULT 6
US-09-430-669-2
; Sequence 2, Application US/09430669
; Patent No. 6329187
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglycanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: 28-Oct-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: 22-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Heiron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: <Unknown>
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-430-669-2

Query Match 38.2%; Score 758; DB 4; Length 553;
Best Local Similarity 45.2%; Pred. No. 1.9e-64;
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNPVRIAGINWFGFETCNVYVHGLRSRYSRMLDQIKSLGYNTRLRPSYDDILKP 72
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPCTQSVKP 63

QY 73 GTPNSINFYQMNQDLOGLSLQVMDKIVAYAGOIGRLIILDRHRPDCSGQSALWYTSV 132
DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEXIIKKAGDLGIFVLLDYHRIICNFIEPLWYTSF 122

QY 133 SEATWISDLQALQRYKGNPTVVGFDLHNEPDP-----ACWCGDPSIDWRL 180
DB 123 SEQDYINTWVEAQRFGKYNWIGADLKNPHESSPAPAAAYTDGSGATWGMGNATDNL 182

QY 181 AAERAGNAVLSVNPMLLIIFVEGVQ-----SYN-GDSYWGNGNLQAGQYPPVVLNVP- 230
DB 183 AAERIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWHNAWGNLGMVRYKYPV--NLPR 240

QY 231 NRVLYSAHDYATSVYPTWFSDP--TFNNMFGIWNKNGYLFNQNIAPVWLGEFGTTL- 287
DB 241 DKVYSPQVYSGVYDQPYF-DPGEFPDNLPEIWHFHYKVLDLGYFVWIGFEGKYG 299

QY 288 --QSTTDQTLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSWNPNSGDTGGILKDDWTTI 348

RESULT 7
US-09-869-197-1
; Sequence 1, Application US/09869197
; Patent No. 6566113
; GENERAL INFORMATION:
; APPLICANT: TAKAYAMA, Masanori
; APPLICANT: UMEDA, Kahoko
; APPLICANT: KOYAMA, No. 6566113uto
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: TAKAYAMA-6
; CURRENT APPLICATION NUMBER: US/09/869,197
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/JP99/07009
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: JP 365237/1998
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii OT3
US-09-869-197-1

Query Match 37.9%; Score 751; DB 4; Length 430;
Best Local Similarity 45.0%; Pred. No. 6.2e-64;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNPVRIAGINWFGFETCNVYVHGLRSRYSRMLDQIKSLGYNTRLRPSYD 67
DB 26 TSGBE-----TPHILFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPCT 79

QY 68 DILKPGTNPNSINFYQMNQDLOGLSLQVMDKIVAYAGOIGRLIILDRHRPDCSGQSALW 127
DB 80 ESKVPGTQPIGID-YSRNPDLRGLDSQIMEXIIKKAGDLGIFVLLDYHRIICNFIEPLW 138

QY 128 YTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPDP-----PACWCGDPS 175
DB 139 YTEDFSEDFINTWIEVAKRFGKYNWIGADLKNPHESSPAPAAAYTDGSGATWGMGNPA 198

QY 176 IDWRLAERAGNAVLSVNPMLLIIFVEGVQSYN-----GDSYWGNGNLQAGQYPPV 226

Query Match 99.6%; Score 1974; DB 1; Length 358;
Best Local Similarity 99.7%; Pred. No. 6.7e-182;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSMLODKISLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSMLODKISLGYNT 60
QY 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180
DB 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180
QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOAGAGQYPVVLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOAGAGQYPVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTWFSDFPTFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTLOSTDTQWLKTLV 300
DB 241 ATSVYPTWFSDFPTFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTLOSTDTQWLKTLV 300
QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTVTDVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTVTDVKDGYLAPIKSSIFDPV 358

RESULT 2

US-08-276-213-3
; Sequence 3, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Laymon, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 99.6%; Score 1974; DB 1; Length 521;
Best Local Similarity 99.7%; Pred. No. 1.2e-181;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSMLODKISLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSMLODKISLGYNT 60
QY 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180
DB 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180
QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOAGAGQYPVVLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOAGAGQYPVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTWFSDFPTFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTLOSTDTQWLKTLV 300
DB 241 ATSVYPTWFSDFPTFPNNMFGIWNKMGYLFNQNIAPVWLGEFGTTLOSTDTQWLKTLV 300
QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTVTDVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDWQTVTDVKDGYLAPIKSSIFDPV 358

RESULT 3

US-08-651-572-2
; Sequence 2, Application US/08651572
; Patent No. 5789228
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-651-572-2

us-09-997-504a-12.ra1

Wed Aug 4 10:02:19 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds
(without alignments)
1260.143 Million cell updates/sec

Title: US-09-997-504A-12
Perfect score: 1982
Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDFV 358
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgm2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgm2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgm2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgm2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	358	1	US-08-604-913B-11
2	1974	99.6	521	1	US-08-276-213-3
3	758	38.2	553	1	US-08-651-572-2
4	758	38.2	553	3	US-09-066-544-2
5	758	38.2	553	3	US-08-951-086-2
6	758	38.2	553	4	US-09-430-669-2
7	751	37.9	430	4	US-08-869-197-1
8	751	37.9	458	4	US-09-869-197-5
9	617	31.1	616	3	US-09-136-574A-47
10	617	31.1	1426	3	US-09-136-574A-43
11	205	10.3	38	1	US-08-276-213-2
12	178.5	9.0	377	1	US-08-525-697-2
13	141	7.1	400	4	US-09-797-464A-4
14	137	6.9	348	4	US-09-797-464A-7
15	136	6.9	395	4	US-09-797-464A-2
16	134	6.8	363	4	US-09-797-464A-11
17	133.5	6.7	663	4	US-09-134-078-61
18	133.5	6.7	680	4	US-09-134-078-25
19	132	6.7	476	4	US-09-339-159B-4
20	132	6.7	490	4	US-09-339-159B-2
21	131	6.6	24	1	US-08-276-213-1
22	127	6.4	331	4	US-09-339-159B-12
23	126	6.4	468	4	US-09-485-648-4
24	126	6.4	468	4	US-09-503-565-4
25	126	6.4	468	4	US-09-485-649-4
26	126	6.4	468	4	US-09-339-159B-8
27	126	6.4	493	4	US-09-485-648-2

28	126	6.4	493	4	US-09-503-565-2
29	126	6.4	493	4	US-09-485-649-2
30	126	6.4	493	4	US-09-339-159B-6
31	124	6.3	320	4	US-09-339-159B-22
32	114	5.8	360	4	US-09-134-078-27
33	113	5.7	461	1	US-08-672-571A-3
34	113	5.7	490	1	US-08-672-571A-1
35	112.5	5.7	429	1	US-08-745-977-4
36	112.5	5.7	429	3	US-08-040-699A-4
37	111	5.6	327	2	US-08-169-948B-16
38	111	5.6	327	3	US-08-448-873-16
39	111	5.6	327	3	US-08-382-452D-16
40	111	5.6	327	4	US-08-507-362A-8
41	111	5.6	327	4	US-08-916-494A-16
42	109.5	5.5	418	3	US-09-254-733-5
43	109.5	5.5	430	2	US-08-924-440-2
44	108	5.4	317	2	US-09-066-075-2
45	108	5.4	317	2	US-08-518-615A-2

ALIGNMENTS

RESULT 1
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.
; APPLICANT: Himmel, Michael E.
; APPLICANT: Baker, John O.
; APPLICANT: Chou, Yat-Chen
; TITLE OF INVENTION: METHOD FOR INCREASING
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASC II (DOS) text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/604,913B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,213
; FILING DATE: 15-07-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Edna M. O'Connor
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: 95-56
; TELEPHONE: 303/384-7573
; TELEFAX: 303/384-7499
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; FEATURE:
; NAME/KEY: E1-CAT
; US-08-604-913B-11

Search completed: August 2, 2004, 16:34:25
Job time : 49.6667 secs

DR WPI; 2001-126236/14.

XXXXXXXXXXXX

XXXXXXXXXXXX

KW emulsion stability; temperature resistance; pseudoplasticity; amylase;
 KW cellulase; extracellular protease; intracellular protease;
 KW Glucose dehydrogenase; enzyme.
 XX Xanthomonas campestris.
 XX US2003036176-A1.
 XX 20-FEB-2003.
 XX 10-AUG-2001; 2001US-00927827.
 XX 28-MAR-2001; 2001US-0279493P.
 XX (BOWE/) BOWER S G.
 XX (RAMS/) RAMSEIER T M.
 XX Bower SG, Ramseier TM;
 XX WPI; 2003-625389/59.
 XX N-PSDB; ADD24896.
 XX New transformed cell or organism having reduced or enhanced activity of
 PT at least one protein useful for producing xanthan gum, which are useful
 PT for providing formulations and properties.
 XX Claim 1; SEQ ID NO 50; 135pp; English.
 XX The present invention relates to polypeptide and polynucleotide sequences
 CC from Xanthomonas campestris which may be used for activity reduction or
 CC enhancement using directed genetic engineering. A transformed cell or
 CC organism having reduced or enhanced activity of at least one such protein
 CC e.g. galactomannanase can be generated by disrupting the gene encoding
 CC the protein. The activity of the protein is reduced by the presence of an
 CC antisense nucleic acid sequence. The nucleic acid sequence of the gene
 CC encoding the protein is a recombinant sequence having at least one
 CC mutation as compared to the wild-type gene encoding the protein. The
 CC transgenic cell or microorganism are useful for producing xanthan gum,
 CC which are useful for providing formulations and properties, such as long-
 CC term suspension and emulsion stability in alkaline, acid, and salt
 CC solutions, temperature resistance, and pseudoplasticity. The present
 CC sequence represents an enzyme relating to the present invention.
 XX Sequence 535 AA;
 SQ
 Query Match 39.6%; Score 785.5; DB 7; Length 535;
 Best Local Similarity 46.8%; Pred. No. 7.9e-66;
 Matches 160; Conservative 59; Mismatches 105; Indels 19; Gaps 9;
 QY 5 YHWTGREITLDANNVPRIAGINWGFETCNVYVHGLRSDYRSMDDQIKSLGYNTIRLP 64
 DB 77 YSINNSRQIVDDSGKVQLKGVNVGFETCNVHGLWARKMDIVQMGGJGFNAVRLP 136
 QY 65 YSDIILKPGTTPMNSINFYQNNQDGLTSLQWMDKIVAYAGQIGLRILDRHRPDCSGS 124
 DB 137 FCPATLRSMTPASID-YERNADQLTSLQIDKVIABFNARGMYVLLDHTPDGAGIS 195
 QY 125 ALWYTSVSEATWISDLQALQAYKGNPTVGFGLHNEPHDPCMGCGPDSIDWRLAER 184
 DB 196 ELYTSGSYTEAQLADLRVFNARYKNVPYVGLDLKNEPHGATWGTGNAATDNKAER 255
 QY 185 AGNAVLSVNPNNLIFVEGVQ-----SYNGDSYMWGNLQAGQYPVVLPVNP-NRLVYSAH 238
 DB 256 GSAALVAVAPKWLIAVEGITDNPVCSITNG-GIFWGNLQPLACTP--LNIPANRLILAPH 312
 QY 239 DYATSVYPTWFRSDPTFPNNMPCGINKNNGWLFENQNIAPVWLGEGFTTL--OSTTDQWL 296
 DB 313 VYGPDFVQSIFNDSNFPNMPAIWHRFGQFAGTH--ALLGEFDCKYGEGBDARDKTWQ 370
 QY 297 KTLVQLRPTAQYGADSFQWTSWNPDSGDTGKILKDDQTV 339
 DB 371 DALVKYLR---SKGIN--QGFWYWNPNFNSGDTGILRDDWTSV 408

RESULT 9

AAW34999
 ID AAW34999 standard; protein; 628 AA.
 XX
 AC AAW34999;
 XX
 XX 27-AUG-2003 (revised)
 DT 21-MAY-1998 (first entry)
 XX
 DE Archaeobacterium AEP11a endoglucanase.
 XX
 XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
 KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
 KW thermostable enzyme; thermophilic; glycosidase.
 XX
 OS archaean.
 OS Unidentified...
 XX
 XX WO9744361-A1.
 PN
 XX 27-NOV-1997.
 PD
 XX 22-MAY-1997; 97WO-US008793.
 PF
 XX 22-MAY-1996; 96US-00651572.
 PR
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
 PA
 XX Lam DE, Mathur EJ;
 PI
 XX WPI; 1998-018435/02.
 DR
 XX N-PSDB; AAT94207.
 XX
 PT Endo:glucanase(s), preferably form archaeal bacterium, AEP11a - useful
 PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
 PT bonds in cellulose.
 XX
 PS Claim 1; Fig 10; 164pp; English.
 XX
 CC This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone
 CC 630P2), a hydrothermal vent isolate. The endoglucanase is capable of
 CC degrading carboxymethylcellulose and of hydrolysing the beta-1,4-
 CC glycosidic bonds in cellulose. It has homology to another endoglucanase
 CC (see AAW34985) of archaeobacterium AEP11a. It can be produced from native
 CC cells or from recombinant host cells, especially prokaryotic host cells
 CC transformed with a plasmid or virus-derived vector including the
 CC endoglucanase DNA (see AAT94207). 24 Endoglucanases (see AAW34986-W35008)
 CC are claimed. They can be used to degrade cellulose for the conversion of
 CC plant biomass into fuels and chemicals, for use in detergents, textiles,
 CC animal feed, waste treatment, and in the fruit juice and brewing
 CC industries for the clarification and extraction of juices. (Updated on 27
 CC -AUG-2003 to correct OS field.)
 XX
 SQ Sequence 628 AA;
 Query Match 38.3%; Score 759; DB 2; Length 628;
 Best Local Similarity 45.5%; Pred. No. 3.4e-63;
 Matches 161; Conservative 48; Mismatches 109; Indels 36; Gaps 11;
 QY 13 ILDANNVPRIAGINWGFETCNVYVHGLRSDYRSMDDQIKSLGYNTIRLPYSDIILK 72
 DB 4 VATGEETPIHLFGVNVGFETPNVYVHGLSRNWMEDMLLQIKSLGFNAIRLPCTQSVKP 63
 QY 73 GTMPNSINFYQNNQDGLTSLQWMDKIVAYAGQIGLRILDRHRPDCSGSALWYTSV 132
 DB 64 GTMPTAID-YAKNPDQLGDSVQIMKIKKAGDLGIFVLLDYHICGNFIEPLWYDTSF 122
 QY 133 SEATWISDLQALQAYKGNPTVGFGLHNEPHD-----ACWCGDPSIDWRL 180
 DB 123 SEQDYINTWVEVAQRFQKYNVNVIGADLKNEPHSSPPAPAYTDGSGATWGNNAATDNL 182
 QY 181 AERAGNAVLSVNPNNLIFVEGVQ-----SYN-GDSYMWGNLQAGQYPVVLPVNP- 230

Db 282 ATSVPTQWFSPTFPNNPFGIWNKNGVLFNQNIAPVWLGEFGTTLQSTTDQTLKTLV 341
 QY 301 QYLRTAQYGADSFQWTFWNNPDSGDTGGILKDDQWQVTVKDGYLAPIKSSIFDPV 358
 Db 342 QYLRTAQYGADSFQWTFWNNPDSGDTGGILKDDQWQVTVKDGYLAPIKSSIFDPV 399

RESULT 5
 AAY69508
 ID AAY69508 standard; protein; 562 AA.
 XX
 AC AAY69508;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE Acidothermus cellulolyticus E1 endoglucanase.
 XX
 KW E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;
 XX beta-1,4-endoglucanase; endocellulase; thermostable.
 KW
 OS Acidothermus cellulolyticus.
 XX

Key	Location/Qualifiers
Peptide	1..41
Peptide	/note= "Putative signal peptide"
Peptide	14..41
Peptide	/note= "Putative signal peptide (alternative)"
Domain	42..404
Domain	/note= "Catalytic domain"
Region	405..460
Region	/note= "Linker region"
Domain	461..562
Domain	/note= "Cellulose binding domain (CBD)"

CA2226898-Al.
 25-SEP-1999.
 25-MAR-1998; 98CA-02226898.
 25-MAR-1998; 98CA-02226898.
 (MIDE) MIDWEST RES INST.
 Laymon RA, Adney WS, Thomas SR, Himmel ME;
 WPI; 2000-087663/08.
 N-PSDB; AA255924.

Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful for labeling or modifying a cellulose and for purifying or immobilizing a binding domain fusion protein to cellulose.

Claim 1; Fig 2; 85pp; English.

This sequence represents the Acidothermus cellulolyticus E1 endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which encode it are specifically claimed. The CBD is believed to be roughly wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge tip is inserted between the microfibrils of the cellulose fibre, disrupting the crystalline structure, and making the cellulose linkages more accessible to the catalytic domain of the E1 endoglucanase. The E1 endoglucanase CBD is useful in labelling or modifying the surface of cellulose or other polysaccharides. Such modified cellulose can then be used in textile, pulp, paper, chemical and pharmaceutical industries. CBDs can be used in affinity purification of CBD-fusion proteins, and can also be used to immobilise the CBD-fusion proteins to a cellulose support. CBD-fusion proteins can be used to modify the chemical or physical properties of a cellulose or polysaccharide matrix column and to modify (e.g. roughen or disrupt) a cellulose or polysaccharide fibre. The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDS from non-thermophilic organisms
 XX
 SQ Sequence 562 AA;
 Query Match 99.4%; Score 1970; DB 3; Length 562;
 Best Local Similarity 99.4%; Pred. No. 7.1e-179;
 Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRMLDQIKSLGYNT	60
Db	42	AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRMLDQIKSLGYNT	101
QY	61	IRLPYSDDILKPGTNPNSINFYQWQDLOGLTSLQWQDKIVAYAGQIGRIILDRHRPDC	120
Db	102	IRLPYSDDILKPGTNPNSINFYQWQDLOGLTSLQWQDKIVAYAGQIGRIILDRHRPDC	161
QY	121	SGQSALWYTSVSEATWISDLOALQRYKGNPVVGFDLHNEPHDPACWCGDPSIDWRL	180
Db	162	SGQSALWYTSVSEATWISDLOALQRYKGNPVVGFDLHNEPHDPACWCGDPSIDWRL	221
QY	181	AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYFVVLNVNRLVYSAHDY	240
Db	222	AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYFVVLNVNRLVYSAHDY	281
QY	241	ATSVPTQWFSPTFPNNPFGIWNKNGVLFNQNIAPVWLGEFGTTLQSTTDQTLKTLV	300
Db	282	ATSVPTQWFSPTFPNNPFGIWNKNGVLFNQNIAPVWLGEFGTTLQSTTDQTLKTLV	341
QY	301	QYLRTAQYGADSFQWTFWNNPDSGDTGGILKDDQWQVTVKDGYLAPIKSSIFDPV	358
Db	342	QYLRTAQYGADSFQWTFWNNPDSGDTGGILKDDQWQVTVKDGYLAPIKSSIFDPV	399

RESULT 6
 AAB48787
 ID AAB48787 standard; protein; 521 AA.
 XX
 AC AAB48787;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.
 XX
 KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation; mutant; muten.
 XX
 OS Acidothermus cellulolyticus.
 OS Synthetic.
 XX
 PN WO200070031-Al.
 XX
 PD 23-NOV-2000.
 XX
 PF 19-MAY-2000; 2000WO-US013971.
 XX
 PR 19-MAY-1999; 99US-0134925P.
 XX
 PA (MIDE) MIDWEST RES INST.
 XX
 PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 PI Decker SR;
 XX
 DR WPI; 2001-061226/07.
 XX
 PT Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.
 PT
 PS Claim 17; Page 25-27; 30pp; English.
 XX
 CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic

XX 15-JUL-1994; 94US-00276213.
 XX (MIDE) MIDWEST RES INST.
 XX Thomas SR, Laymon RA, Himmel ME;
 XX WPI; 1996-105843/11.
 XX N-PSDB; AAT12337.
 XX New isolated DNA encoding endo:glucanase - obtd from Acidothermus
 XX cellulolyticus, used for prodn of the enzyme for use in cellulose
 XX hydrolysis.
 XX Claim 1; Page 22; 34pp; English.
 XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful for
 XX hydrolysing cellulosic biomass to sugars for simultaneous or subsequent
 XX fermentation to ethanol. It shows optimal activity at 83 deg C. The amino
 XX acid sequence of the mature enzyme was deduced from an E1 endoglucanase
 XX gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides
 XX were also identified (see also AAR89928-29). Cloning of this gene allows
 XX large-scale, low-cost prodn. of recombinant E1 endoglucanase, using pref.
 XX Saccharomyces, Zymomonas or E.coli hosts
 XX Sequence 521 AA;
 XX
 XX Query Match 99.6%; Score 1974; DB 2; Length 521;
 XX Best Local Similarity 99.7%; Pred. No. 2.6e-179;
 XX Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60
 DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60
 QY 61 IRLPYSDILKPGTNPNSINFYQNMNQLQGLTSLOWMDKIVAYAGQIGRLILDRHRPDC 120
 DB 61 IRLPYSDILKPGTNPNSINFYQNMNQLQGLTSLOWMDKIVAYAGQIGRLILDRHRPDC 120
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
 DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
 QY 181 AAERAGNAVLSPNLLIFVEGVSYNGDSYVWGGNLQAGQYPPVNLNPNRLVYSAHDY 240
 DB 181 AAERAGNAVLSPNLLIFVEGVSYNGDSYVWGGNLQAGQYPPVNLNPNRLVYSAHDY 240
 QY 241 ATSVYPQTWFSDDPTFPNNMPCINWKNWGYLFNQNIAPVWLGEFTTLQSTTDQTLWTLV 300
 DB 241 ATSVYPQTWFSDDPTFPNNMPCINWKNWGYLFNQNIAPVWLGEFTTLQSTTDQTLWTLV 300
 QY 301 QYLRPTAQYGADSFQWTFWSNPDGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
 DB 301 QYLRPTAQYGADSFQWTFWSNPDGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
 RESULT 4
 AAU79549
 ID - AAU79549 standard; protein; 562 AA.
 XX
 XX AAU79549;
 XX
 XX 24-SEP-2002 (first entry)
 XX
 XX A. cellulolyticus cellulase E1 beta-1,4-endoglucanase precursor.
 DE
 XX Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;
 XX fermentable sugar; ethanol; fermentation; silage; feed; fuel;
 XX industrial chemical; biodegradation; chloroaromatic;
 XX environmental pollutant; E1 beta-1,4-endoglucanase; el.
 XX Acidothermus cellulolyticus.
 OS
 XX

XX Key Location/Qualifiers
 FT 1. 41
 FT /label= Leader_sequence
 FT Protein 42. 562
 FT /label= Mature_cellulase
 XX WO200234926-A2.
 XX
 XX 02-MAY-2002.
 XX
 XX 18-OCT-2001; 2001WO-US032538.
 XX
 XX 20-OCT-2000; 2000US-0242408P.
 XX
 XX (UNWS) UNIV MICHIGAN STATE.
 XX
 XX Sticklen WB, Dale BE, Maqbool S;
 XX
 XX WPI; 2002-489947/52.
 XX N-PSDB; ABR86729.
 XX
 XX Producing transgenic plants which after harvest degrade lignin and
 XX cellulose to fermentable sugars, by mating transgenic plant comprising
 XX cell encoding cellulase with transgenic plant comprising DNA encoding
 XX ligninase.
 XX
 XX Disclosure; Page 96-99; 126pp; English.
 XX
 XX The invention discloses the production of a transgenic plant which
 XX degrades lignocellulose when the plant is ground. It comprises the
 XX production of the transgenic plant including cellulase and ligninase by
 XX mating a transgenic plant, containing a DNA encoding a cellulase, and a
 XX transgenic plant, containing a DNA encoding a ligninase, where both genes
 XX are operably linked to a nucleotide sequence encoding a signal peptide
 XX which targets the fusion protein to an organelle of the plant,
 XX particularly chloroplasts. The method is useful for producing a
 XX transgenic plant (e.g. maize) which degrades lignocellulose when the
 XX plant is ground to produce a plant material. This material is useful for
 XX converting lignocellulose, in a plant material, to fermentable sugars
 XX which are then fermented to ethanol. The transgenic plants also provide a
 XX plentiful and inexpensive source of fungal or bacterial cellulases and
 XX ligninases which can be used in the production of ethanol. They can also
 XX be used for pre-treating silage to increase the energy value of
 XX lignocellulosic feeds for cows and other ruminant animals, pre-treating
 XX lignocellulosic biomass for fermentative conversion to fuels and
 XX industrial chemicals, and biodegradation of chloroaromatic environmental
 XX pollutants. The protein sequence presented is the A. cellulolyticus
 XX cellulase E1 beta-1,4-endoglucanase precursor
 XX
 XX Sequence 562 AA;
 XX
 XX Query Match 99.6%; Score 1974; DB 5; Length 562;
 XX Best Local Similarity 99.7%; Pred. No. 2.9e-179;
 XX Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60
 DB 42 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 101
 QY 61 IRLPYSDILKPGTNPNSINFYQNMNQLQGLTSLOWMDKIVAYAGQIGRLILDRHRPDC 120
 DB 102 IRLPYSDILKPGTNPNSINFYQNMNQLQGLTSLOWMDKIVAYAGQIGRLILDRHRPDC 161
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
 DB 162 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 221
 QY 181 AAERAGNAVLSPNLLIFVEGVSYNGDSYVWGGNLQAGQYPPVNLNPNRLVYSAHDY 240
 DB 222 AAERAGNAVLSPNLLIFVEGVSYNGDSYVWGGNLQAGQYPPVNLNPNRLVYSAHDY 281
 QY 241 ATSVYPQTWFSDDPTFPNNMPCINWKNWGYLFNQNIAPVWLGEFTTLQSTTDQTLWTLV 300

CC (AA848788) has improved activity with insoluble substrates, and the W42R
CC (AA848786) and Y82R (AA848787) mutants have improved activity with
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant
XX
SQ Sequence 521 AA;

Query Match 100.0%; Score 1982; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 4.5e-180;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNT 60
Db 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNT 60
QY 61 IRLPYSDDLKPGTTPNSINFYOMNQDLOGLTSLQVMDKIVAYAGIGLRIILDRHRPDC 120
Db 61 IRLPYSDDLKPGTTPNSINFYOMNQDLOGLTSLQVMDKIVAYAGIGLRIILDRHRPDC 120
QY 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
Db 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGLQAGQYPPVVLNPNRLVYSAHDY 240
Db 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGLQAGQYPPVVLNPNRLVYSAHDY 240
QY 241 ATSVYPTQWFSDDTFFNNMGIWKNWGYLFNQNIAPIWLGEGFTTLOSTTDQWLKTLV 300
Db 241 ATSVYPTQWFSDDTFFNNMGIWKNWGYLFNQNIAPIWLGEGFTTLOSTTDQWLKTLV 300
QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358
Db 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2

AAW39262
ID AAW39262 standard; protein; 358 AA.

AC AAW39262;

XX 14-MAY-1998 (first entry)

DT A. cellulolyticus E1-CAT translated region.

DE E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;
KW truncated; cellulose hydrolysis; biomass conversion.

XX Acidothermus cellulolyticus.

OS US5712142-A.

PN 27-JAN-1998.

XX 22-FEB-1996; 96US-00604913.

PR 26-SEP-1989; 89US-00412434.

PR 27-JAN-1992; 92US-00826089.

PR 21-SEP-1993; 93US-00125115.

PR 15-JUL-1994; 94US-00276213.

PA (MIDE) MIDWEST RES INST.

XX Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;

XX WPI; 1998-119985/11.

XX N-PSDB; AAV09659.

XX DNA encoding truncated form of E1 endo-glucanase from Acidothermus

PT cellulolytic - and related vectors and transformed cells, expressing
PT only catalytic domain, is used for biomass conversion and has better heat
PT stability than complete enzyme.

XX Claim 2; Fig 5; 19pp; English.

XX This sequence represents a novel Acidothermus cellulolyticus endoglucanase
CC E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the
CC enzyme, without its peptide linker or cellulose binding domain (CBD) and
CC the truncated enzyme expressed by is used for hydrolysis of cellulose
CC (biomass conversion). Compared with full-length E1, the truncated enzyme
CC has better heat stability and higher temperature of maximum activity

XX Sequence 358 AA;

Query Match 99.6%; Score 1974; DB 2; Length 358;

Best Local Similarity 99.7%; Pred. No. 1.5e-179;

Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNT 60

Db 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNT 60

QY 61 IRLPYSDDLKPGTTPNSINFYOMNQDLOGLTSLQVMDKIVAYAGIGLRIILDRHRPDC 120

Db 61 IRLPYSDDLKPGTTPNSINFYOMNQDLOGLTSLQVMDKIVAYAGIGLRIILDRHRPDC 120

QY 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

Db 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGLQAGQYPPVVLNPNRLVYSAHDY 240

Db 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGLQAGQYPPVVLNPNRLVYSAHDY 240

QY 241 ATSVYPTQWFSDDTFFNNMGIWKNWGYLFNQNIAPIWLGEGFTTLOSTTDQWLKTLV 300

Db 241 ATSVYPTQWFSDDTFFNNMGIWKNWGYLFNQNIAPIWLGEGFTTLOSTTDQWLKTLV 300

QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

Db 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3

AA889927
ID AA889927 standard; protein; 521 AA.

AC AA889927;

XX 08-OCT-1996 (first entry)

DT A. cellulolyticus E1 endoglucanase.

DE E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.

XX Acidothermus cellulolyticus.

OS Key Location/Qualifiers

FT Domain 1..363

FT Domain /label= Catalytic-domain

FT Domain 364..417

FT Domain /label= Linker

FT Domain /note= "proline/serine/threonine-rich linker domain

FT Domain 418..521

FT Domain /label= Cellulose-binding_domain

XX WO9602551-A1.

XX 01-FEB-1996.

XX 14-JUL-1995; 95WO-US008868.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:04 ; Search time 47.6667 Seconds
(without alignments)
2122.070 Million cell updates/sec

Title: US-09-997-504A-12
Perfect score: 1982
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1980s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1982	100.0	521	4 AAB48786	Aab48786 Acidother
2	1974	99.6	358	2 AAW39262	Aaw39262 A. cellul
3	1974	99.6	521	2 AAR89927	Aar89927 A. cellul
4	1974	99.6	562	5 AAU79549	Aau79549 A. cellul
5	1970	99.4	562	3 AAY69508	Aay69508 Acidother
6	1965	99.1	521	4 AAB48787	Aab48787 Acidother
7	1964	99.1	521	4 AAB48788	Aab48788 Acidother
8	785.5	39.6	535	7 ADD24922	Add24922 Xanthomon
9	759	38.3	628	2 AAW34999	Aaw34999 Archaeabac
10	759	38.3	841	2 AAW34985	Aaw34985 Archaeabac
11	758	38.2	553	5 ABG70759	Abg70759 T. mariti
12	751	37.9	430	3 AAB10344	Aab10344 P. horiko
13	751	37.9	458	3 AAB10345	Aab10345 P. horiko
14	694.5	35.0	514	4 AAB96264	Aab96264 Putative
15	617	31.1	616	2 AAY13494	Aay13494 Truncated
16	617	31.1	616	2 AAE16325	Aae16325 Active ce
17	617	31.1	1426	2 AAY13492	Aay13492 Truncated
18	617	31.1	1426	5 AAE16323	Aae16323 Active ce
19	509.5	25.7	425	4 ABB03060	Abb03060 Thermus c
20	281.5	14.2	582	7 ADD24920	Add24920 Xanthomon
21	267	13.5	508	5 ABB92443	Abb92443 Herbicida
22	205	10.3	38	2 AAR89930	Aar89930 A. cellul
23	205	10.3	389	5 AAM49427	Aam49427 Penicilli
24	204.5	10.3	488	5 ABB93510	Abb93510 Herbicida
25	189	9.5	551	5 ABB92442	Abb92442 Herbicida

26	181.5	9.2	762	6 ABP73022	Abp73022 Amino aci
27	179.5	9.1	375	6 ABP73019	Abp73019 Amino aci
28	176	8.9	517	6 ABP99336	Abp99336 Orthosomy
29	154.5	7.8	722	6 ABU24069	Abu24069 Protein e
30	151	7.6	420	6 ABP96833	Abp96833 Candida u
31	149.5	7.5	329	2 AAW35002	Aaw35002 Thermotog
32	146.5	7.4	518	4 ABB06928	Abb06928 Micromono
33	146	7.4	666	2 AAW34992	Aaw34992 Thermotog
34	145.5	7.3	406	2 AAR08199	Aar08199 Neutral c
35	142	7.2	431	3 AAG42172	Aag42172 Arabidops
36	142	7.2	431	3 AAG27501	Aag27501 Arabidops
37	142	7.2	431	5 ABB93972	Abb93972 Herbicida
38	142	7.2	442	3 AAG42171	Aag42171 Arabidops
39	142	7.2	443	3 AAG27500	Aag27500 Arabidops
40	141	7.1	335	7 ADD24924	Add24924 Ralstonia
41	141	7.1	400	4 AAE09785	Aae09785 Paenibaci
42	138	7.0	375	3 AAG27502	Aag27502 Arabidops
43	138	7.0	375	3 AAG42173	Aag42173 Arabidops
44	138	7.0	562	4 AAU42094	Aau42094 Propionib
45	138	7.0	562	6 AEM38613	Aem38613 Propionib

ALIGNMENTS

RESULT 1
AAB48786
ID AAB48786 standard; protein; 521 AA.
XX

AC AAB48786;

XX 09-MAR-2001 (first entry)

XX Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.

XX E1 endoglucanase; glycosyl hydrolase; soluble substrate;

KW cellulose hydrolysis; ethanol production; fermentation; mutant; muten.

XX Acidothermus cellulolyticus.

OS Synthetic.

XX WO2000070031-A1.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US013971.

XX 19-MAY-1999; 99US-0134925P.

XX (MIDE) MIDWEST RES INST.

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

PI Decker SR;

XX WPI; 2001-061226/07.

PT Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.

PS Claim 16; Page 22-24; 30pp; English.

XX The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic activity comprises replacing an active site-associated amino glycosyl-stabilising amino acid with an amino acid that does not strongly bind a disaccharide product in the active site. Conversely, the method for making a glycosyl hydrolase with increased soluble substrate catalytic activity comprises replacing a hydrophobic substrate-binding amino acid with a positively charged residue. The invention also discloses mutants of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788) produced according to the method of the invention. The Y245G mutant

QY 219 GAGQEVVLNPNRLVSAHDYATSVGPOTWFSDFTEFNNM-PGIWNKNGVLEFNQIAP 277
 Db 440 GVADHPVISAADKLVYSDVHDGPIYMQPWFKKDFDINTLYEECYFNWYIYEQNIAP 499
 QY 278 VMLGFGTTTQSTDTQTLTKTLVQVLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQ 337
 Db 500 MLIGEWGKLIENNRKWLCECLATEI-----AEKKLHHTFWAFNPSADTGGMLDWDK 553
 QY 338 TVDTVKDGYLAP 349
 Db 554 TVDEEKYAIIVP 565

RESULT 14
 Q9EYQ0 PRELIMINARY; PRT; 534 AA.
 AC Q9EYQ0;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Cellulase Cel5-N.
 GN CELN.
 OS Clostridium cellulolyticum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Belaich A., Gaudin C., Gal L.;
 RT "A gene cluster of cellulases of Clostridium cellulolyticum";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF316823; AAG45162.1; --
 DR HSSP; P54583; 1ECE.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR02048; EF-hand.
 DR InterPro; IPR001547; Glyco_Hydro_5.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR PROSITE; PS00018; EF HAND; 2.
 DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
 DR SEQUENCE 534 AA; 59671 MW; 980785F0971E5A59 CRC64;

Query Match 29.8%; Score 591.5; DB 2; Length 534;
 Best Local Similarity 32.8%; Pred. No. 1.1e-37;
 Matches 133; Conservative 60; Mismatches 144; Indels 69; Gaps 13;

QY 6 W-HTSGREILDANNVVRVRIAGINWFGFCNYYVHGLWSRDSYRSMIDQIKSYNTIRLP 64
 Db 38 WLHCVGDKIYDMRGREVLWTGANWFGNCSNVFHGAW-YDVKNILTSVADRGIGLLRVP 96
 QY 65 YSDIIL-----KPG-----TPNSINFYQMNDLQ-----LTSQYMDKIVAYAGQI 107
 Db 97 ISTELLYSWMTGKPNKVSSTASNNPPYTVNPFYDPATDGPKNSEIFDIIMKYCKEL 156
 QY 108 GLRIILDRHRPCSGQS---ALWY-----TSVSEATWISDQALAQRYKGNPTVGF 158
 Db 157 GIKVMDIVHSPDANNSGHMYPLAYGLTETTAGMITTDKVIDTLTWLAGYKNDITLAI 216
 QY 159 LHNEPH-----DPACWCGDPSIDWRLAERAGNAVLSVNPNNLLIFVEGVQSY- 206
 Db 217 LKNEPHGKGYTNAAPTDMAKNDNTDENNKVYAAERCSKEILAVNPKLLIMIEGIEQYP 276
 QY 207 -----NGDSY-----WGCNLOGAGQYPVVNLVNP-RLVYSAHDYATSV 244
 Db 277 KTEKGYTFTDPVWAGSAGDAAPWHGGWGNLGRGVKDYPIDLGPLNSQIVYSPHDYGPSV 336
 QY 245 GPOTWF-SDPTFPNNMFGIWNKNGVLEFNQIAPVWLGFGTTTQSTDTQTLTKTLVQYL 303
 Db 337 YNQSWEFDKFTTQTLDDYWDYTWAYIDDQKIAPLLIGEWGGFMDGAKNKNWTLIRDM 396

QY 304 RPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTVDTVKDGYLAP 349
 Db 397 I-----KXRNHTFWCLNPNNSGDTGGILGNDSWTWDEEKYGLLKP 436

RESULT 15
 Q7X2N2 PRELIMINARY; PRT; 616 AA.
 AC Q7X2N2;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Endoglucanase.
 GN CEL5B.
 OS Thermomonospora fusca.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
 OX NCBI_TaxID=2021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM51;
 RA Posta K., Beki E., Kukolya J., Hornok L.;
 RT "Phylogenetic relationships of Tf cel5B, a new endoglucanase encoding
 gene from Thermobifida fusca";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY298814; AAP56348.1; --
 DR SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;

Query Match 29.3%; Score 582; DB 2; Length 616;
 Best Local Similarity 34.3%; Pred. No. 7.4e-37;
 Matches 130; Conservative 62; Mismatches 149; Indels 38; Gaps 13;

QY 1 AGGYW-HTSGREILDANNVVRVRIAGINWFGFCNYYVHGLWSRDSYRSMIDQIKSLGYN 59
 Db 33 SGTADLHTDGNRIVDSAGNEVMTGAGNFGFNTSERMFHGLWAAANTEDITSMAERGIN 92
 QY 60 TIRLPYSDIIL---KPGTM-PNSINFYQMNDLQGLTSLQVMDKIVAYAGQIGLRIILDR 115
 Db 93 MVRVPISTQLLEWNGQAGSPGVNEY-VNPELAGMTLLEVDFYWLQICEYGLKVMLDV 151
 QY 116 H--RPDCSGQ-SALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDP----- 166
 Db 152 HSAEADNSGHYYPVWYKGDITTEDFYAWEWTERYKNDTIVAADIKNEPHGKANETPR 211
 QY 167 ACWCGDPSID-WRLAERAGNAVLSVNPNNLLIFVEGVQSYNGD-----SYW 212
 Db 212 AKWD-GSTIDNFKHVCETAGKRILANPNMILCEGIEIYPKDQGSWSTGDRDYISTW 270
 QY 213 WGCNLOGAGQYPVVNLV-NPRLVYSAHDYATSVGPQWFSDFTEFNNM-PGIWNKNGYL 270
 Db 271 WGCNLRGVADHPVLDGAHQQLVYSPHDYGPSVFEQWFEQWNRQTLTDEWVRPNWLYI 330
 QY 271 PNQIAPWLGFGTTTQSTDTQTLTKTLVQVLRPTAQYAGDSFQWTFWSNPNDSGDTGG 330
 Db 331 HEDDIAPLLIGEWGGFLDGDGNKMWALTALSLIID-----EKWHHTFWALNPNNSGDTGG 384
 QY 331 ILKDDWQTVDTVKDGYLAP 349
 Db 395 LLNYDWTWDEAKYAFLEKP 403

Search completed: August 2, 2004, 16:36:56
 Job time : 45.3333 secs

829 RGVKYPINIGKYNKVVYSPHDSYQOPWY-PGFTKESLLQDCHRPWAYIMEEN 887
 275 IAPVWLGEFTTILQSTTDQWLKTLVQYLRPTAQYGADSPQWTFWSNPDSDGTGGILKD 334
 888 IAPLLIGEWGHLGDADNKKWYLRDYII-----ENHIHTFWCFNANSBGDTGLGVY 941
 335 DWQTVDTVKDGYLAP 349
 942 DFTTDEKYSFLKP 956

RESULT 12
 Q9X3P6 PRELIMINARY; PRT; 1426 AA.
 AC Q9X3P6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CelB.
 GN CelB.
 OS Caldicellulosiruptor sp. Tok7B.1.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldicellulosiruptor.
 OC NCBI_TaxID=80339;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=TOK7B.1;
 RC MEDLINE=20171169; PubMed=10706665;
 RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
 RA Bergquist P.L.;
 RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
 thermophile Caldicellulosiruptor isolate tok7B.1.";
 RL Curr. Microbiol. 40:333-340(2000).
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
 HYDROLASES).
 CC EMBL; AF078737; AAD30364.1; -.
 DR HSP; Q06851; INBC.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001589; Actbind_actnin.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR001000; Glyco_hydro_10.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00942; CEM_3; 3.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00331; Glyco_hydro_10; 1.
 DR PRINTS; PD00134; GLHDLASE10.
 DR ProDom; PD001947; CBD_3; 3.
 DR SMART; SMO0633; Glyco_10; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Glycosidase; Hydrolase.
 SQ SEQUENCE 1426 AA; 157544 MW; 2983FDB85D09A863 CRC64;

Query Match 31.3%; Score 621; DB 2; Length 1426;
 Best Local Similarity 35.5%; Pred. No. 2.1e-39;
 Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

10 GREILDANNVVRAGINWFGFTCTN-VVHGLWSRDYRSMQDKSLGNTIRL 69
 1022 GNKIVDKGFWLTVGNWFGFTCTN-VVHGLWSRDYRSMQDKSLGNTIRL 1081

70 L---KPGTMPN-SINFQMNODLQGLTSLQWMDKIVAYAGIGLRIILDRH--RPDCSGQ 123
 1082 ILNWSKGIYPKPNINY-VNPELGLTSLVDFVVKTCREVGKIMLDIHSKTDAMGH 1140

124 -SALWYTSSEATWISDLQALQRYKNGNTVVGFDLHNEPHD-----ACWCGDPS 175
 1141 IYPVWYTDITPEDIYKACEWITRYKNDTDITVADLKNPEPHGKQWDSVFAKWDNSTDI 1200

176 IDWRLAERAGNAVLSVNPILLIFEGVQSYNGP-----SYWNGNLOQAGQ 222
 1201 NNWYAAETCAKRIILAKNPNMLIVIEGLEAYPKDDVTWTSKSSDYSTWNGNLRGVKK 1260
 223 YPVVL-NVNPRLVTSADYATSVGPQTW-----FSDPTFPNNMPGIMKNNMGYLFNOMIAP 277
 1261 YPINLGQYQNVVYSPHDYGLVYQOPWYFGFTKDTLYND---CWRDNTYIMDNGIAP 1317
 278 VWLGEFTTLOSTTDQWLKTLVQYLRPTAQYGADSPQWTFWSNPDSDGTGGILKDDWQ 337
 1318 LLIGEWGGLDGGDNKWMYLRDYII-----ENHIHTFWCFNANSBGDTGLGVYDFS 1371
 338 TVDTVKDGYLAP 349
 1372 TWDEQYNFLKP 1383

RESULT 13
 Q9L3J2 PRELIMINARY; PRT; 660 AA.
 AC Q9L3J2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 1,4-beta-cellobiohydrolase.
 GN CELO.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1515;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=P7;
 RC Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RA Zverlov V.V., Schwarz W.H.;
 RL EMBL; A0275975; CAB76938.1; -.
 DR HSP; P54583; IECE.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR InterPro; IPR000437; Prok_lipoprot_5.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Hydrolase.
 SQ SEQUENCE 660 AA; 75247 MW; 04A042002C288CFC CRC64;

Query Match 31.2%; Score 620.5; DB 2; Length 660;
 Best Local Similarity 35.8%; Pred. No. 8.2e-40;
 Matches 133; Conservative 66; Mismatches 138; Indels 35; Gaps 10;

6 W-HTSGREILDANNVVRAGINWFGFTCTN-VVHGLWSRDYRSMQDKSLGNTIRL 63
 201 WLRVEGNLIKQAGNTVLTGINWFGFTCTN-VVHGLWSRDYRSMQDKSLGNTIRL 260

64 PYSDIIL---KPGTWPNS--INFQMNODLQGLTSLQWMDKIVAYAGIGLRIILDRH-R 117
 261 PLSAIIILQKNGRVERVSFVNTYE-NPRLDGLSLEILDITINHWKNGKAMIDMESS 319

118 PDCSQSALWYTSSEATWISDLQALQRYKNGNTVVGFDLHNEPH-----DPACWGC 171
 320 TKDSYQENLWYKNDITMBEIEAWKWIIVERYDDDDITVADLKNPEPHGKSGPNIAKWD 379

172 GPSIDWRLAERAGNAVLSVNPILLIFEGVQSYNGDSY-----WVGNGNQ 218
 380 SNDPNNWKAELIAEELAINPNLLIVVEGVEAPMEGYDYNCGEFTTYCNWNGNLR 439

QY 236 SAHDYATSVGPTWFSPTPTNNMPGKNNKNGVLFNQNIAPVWLGEGFTTL--QSTTDQ 293
Db 258 MPHVGPDVYQVYFNFPDNNAAIWDKHFHAKAGYA-MAIGFEGKYGEGDPRDI 316
QY 294 TWLXTLVQYLRPTAQYG-ADSFQWTFWSMNPDSDGTGILKDDWQTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSVDTCGMVGNWDTTPRDDKVK 363

RESULT 11
O24820 PRELIMINARY; PRT; 1000 AA.
AC O24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
RT thermophilic anaerobe strain NA10."
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AB008029; BA222939.1; -.
DR HSP; Q06851; INEC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLYHYDRLASE10.
DR PRODOM; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 31.6%; Score 627.5; DB 2; Length 1000;
Best Local Similarity 35.5%; Pred. No. 4.1e-40;
Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HFSGREILDANNVPYRIAGINWFGFETCNVYVHGLWSRDSMLDQIKSLGYNTIRLP 64
Db 590 WLYVSGNKIVDKGRPVWLTGVNFGYNTGTVDFGVWSCNLKSLAEIANRGNLRLVP 649
QY 65 YSDDIL---KPGTWPN-SINFYQMNQDLGTLISQVMDKIVAYAGQIGLRIILDRH--RP 118
Db 650 ISABELLWNSQGIYPKENINY-VNPELEGKNSLEVPDIVVQTCKEVGLKIMLDIHSIKT 708
QY 119 DCSQ-SALVYTSVSEATWISDLQALAQYKGNPTVVGFDLNEHPDP-----ACWG 170
Db 709 DAMGHIPVWTFDEFTEDFYKACEWITNRYKNDTIIAFDLKNEPGKQWDTTFKWD 768
QY 171 CGDPSIDRWLAERAGNAVLNVPNLLIFVEGVQSYNGD-----SYWGGNLL 217
Db 769 NSTDINNKKYAAETCAKRIININENLLIVIEGTEAYPKDDVTWTSKSSDYSTWGGNLL 828
QY 218 QGAGQYVWLV-NVENRVLVYSAHDYATSVGPQTFWSDFTFNN--MPCGIWKNKNGVLFNQ 274

QY 288 QSTTDQTWLXTLVQYLRPTAQYGADSFQWTFWSMNPDSDGTGILKDDWQTV--DTVK 343
Db 326 EG-LDXWQDAFYKWLKIKKIY-----NFFYCLNFPESGDTGGIFLDDKWTNWEK 375

RESULT 10
O87AH4 PRELIMINARY; PRT; 614 AA.
AC O87AH4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN ENGXA OR 2D181.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuranae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.B. Jr., Sassaki F.T., Sena J.A.D.,
RA de Souza A.L., Truffi D., Teukumo F., Yanai G.M., Zeros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa";
RL J. Bacteriol. 185:1018-1026(2003).
DR EMBL; AB012560; AAC29683.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRINTS; PR01228; EGGSHLL.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 62041 MW; 29EC67E42F4BC3F7 CRC64;

Query Match 32.7%; Score 649.5; DB 16; Length 614;
Best Local Similarity 39.9%; Pred. No. 4.1e-42;
Matches 141; Conservative 58; Mismatches 125; Indels 29; Gaps 12;

QY 5 YHITSGREILDANNVPYRIAGINWFGFETCNVYVHGLWSRDSMLDQIKSLGYNTIRLP 64
Db 26 YSISHGKVDWDDKN-QIQLRGVNWFGFETGDHVNGLWLNWKNKEFITQLQGNGFNAILRP 84
QY 65 YSDDILKPGTWPNSINFYQMNQDLGTLISQVMDKIVAYAGQIGLRIILDRHPRDCSGOS 124
Db 85 PCANLNSNTSPSSID-YSRNPDQLGSLIQLDKVKLSLDRMYVLLDHRPDCSALS 143
QY 125 ALMYTSVSEATWISDLQALAQYKGNPTVVGFDLNEHPDPACGCGPSIDRWLAER 184
Db 144 ELWYTDSEKQWIDDLRFVAHRYANVHVGILEVKNENPHGRITWTGTPKTDWNTAVEH 203
QY 185 AGNAVLNVPNLLIFVEGVQSYNGDS-----YWGNGNLCAGQYPPVLANP-NRLVY 235
Db 204 AAAAILAEPKWLIGVEGI-----GENPICSSITGHFWGENLEPMDCTP--LKVPADHILL 257

Db	377	TIDTKKALVQP	388	
Db	Q9PF60	PRELIMINARY;	PRT;	592 AA.
AC	Q9PF60	(TrEMBLrel. 15, Created)		
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2000	(TrEMBLrel. 25, Last annotation update)		
DE	Endo-1,4-beta-glucanase.			
GN	XP0818.			
OS	Xylella fastidiosa.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
OC	Xanthomonadaceae; Xylella.			
OX	NCBI_TaxID=2371;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=945C.			
RX	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G.; Reinach F.C.; Arruda P.; Abreu F.A.; Acencio M.,			
RA	Alvarenga R.; Alves L.M.C.; Araya J.E.; Bordin S.; Bove J.M.; Briones C.S.,			
RA	Barros M.H.; Bonaccorsi E.D.; Bordin S.; Bove J.M.; Briones C.S.,			
RA	Bueno M.R.P.; Camargo A.A.; Camargo L.E.A.; Carraro D.M.; Carrer H.,			
RA	Colauco N.B.; Colombo C.; Costa F.P.; Costa M.C.R.; Costa-Netto C.M.,			
RA	Coutinho L.L.; Cristofani M.; Dias-Netto E.; Docena C.; El-Dorri H.,			
RA	Faccinani A.P.; Ferreira A.J.S.; Ferreira V.C.A.; Ferro J.A.,			
RA	Fraga J.S.; Franca S.C.; Franco M.C.; Fromme M.; Furlan L.R.,			
RA	Garnier M.; Goldman G.H.; Goldman M.H.S.; Gomes S.L.; Gruber A.,			
RA	Ho P.L.; Hoheisel J.D.; Junqueira M.L.; Kemper E.L.; Kitajima J.P.,			
RA	Kleger J.E.; Kuramae E.E.; Laigret F.; Lambais M.R.; Leite L.C.C.,			
RA	Lenos E.G.M.; Lemos M.V.F.; Lopes S.A.; Lopes C.R.; Machado J.A.,			
RA	Machado M.A.; Madeira A.M.B.N.; Madeira H.M.F.; Marino C.L.,			
RA	Marques M.V.; Miranda E.A.L.; Martins E.M.F.; Matsukuma A.Y.,			
RA	Mench C.F.M.; Miracca E.C.; Miyaki C.Y.; Monteiro-Vitorello C.B.,			
RA	Moore D.H.; Nagai M.A.; Nascimento A.L.T.O.; Netto L.B.S.,			
RA	Nhani A.Jr.; Nobrega F.G.; Nunes L.R.; Oliveira M.A.,			
RA	de Oliveira M.C.; de Oliveira R.C.; Palmieri D.A.; Paris A.,			
RA	de Oliveira B.R.; Pereira G.A.G.; Pereira H.A. Jr.; Pesquero J.B.,			
RA	Quaggio B.B.; Roberto P.G.; Rodrigues V.; de Rosa A.J.M.,			
RA	de Rosa V.E. Jr.; de Sa R.G.; Santelli R.V.; Sawasaki H.E.,			
RA	da Silva A.C.R.; da Silva A.M.; da Silva F.R.; Silva W.A. Jr.,			
RA	da Silveira J.F.; Silvestri M.L.Z.; Silveira W.J.; de Souza A.A.,			
RA	de Souza A.P.; Terenzi M.F.; Truffi D.; Tsai S.M.; Tsuchioka M.H.,			
RA	Vallada H.; Van Sluys M.A.; Verjovski-Almeida S.; Vettore A.L.,			
RA	Zago M.A.; Zatz M.; Meidanis J.; Setubal J.C.			
RT	"The genome sequence of the plant pathogen Xylella fastidiosa."			
RL	Nature 406:151-159(2000).			
RL	EMBL; AF003921; AAF83628.1; ..			
DR	PIR; P54583; E82759.			
DR	HSSP; P54583; 1ECE.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001919; Bac celose-bind.			
DR	InterPro; IPR008965; Cellul bind.			
DR	InterPro; IPR001547; Glyco_Hydro_5.			
DR	Pfam; PF00553; CBM_2; 1.			
DR	Pfam; PF00150; cellulase; 1.			
DR	SMART; SM00637; CBD II; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Complete proteome.			
QY	SEQUENCE 592 AA; 59967 MW; 9846DA4E3B5C93E CRC64;			
Query Match	34.6%; Score 687.5; DB 16; Length 592;			
Best Local Similarity	42.6%; Pred. No. 4.3e-45;			
Matches 146; Conservative	50; Mismatches 128; Indels 19; Gaps -10;			
QY	5 YWHTSGREILDANNVPIVRIAGINWFGPFCNTVYVHGLWSRDRYRMLDQIKSLGYNTRILP			64
Db	26 YSIHGKVIDDKGN-QIQLKGSFGFETTHVHGLWTRNWKFFIQLQSMGLNAVRLP			84
QY	65 YSDDLKPGTWNPSINFQNMQLQGLTSLQVMDKIVAYAGIQLRILDRHPCDSCQS			124
Db	85 FCPASLNSNTSPSSID-YNRPNDLQGLSLSLQIMDKVVKELSDRGIIYVMDHHTPDCAAIS			143
QY	125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWELAAER			184
Db	144 ELWYTDYSSEKOWIDDLRFVAVRYANVPGVIGLDVKNRPHGRATWGTGDKXTDNTAVEH			203
QY	185 AGNAVLVSNPNTLIFVEGVQSYNGDS---YWMGNLQAGAGQYFVNLNVP-NRLVYSAHD			239
Db	204 AAAAILEAAPKWLIGVEGIGENPSCSSTIGHFWGLENLEPMDCTP--LKVPADHLLILPHV			261
QY	240 YATSVGPOTWFSDDTFFNNMFGIKNWKGWYLFNQNIAPVWLGEETTL--QSTTDDQWLK			297
Db	262 YGPDVYVQYFNSPDPFNNMAIWDKHFHFAKAGYA-MAIGFEGGKYGEGRDRAWQN			320
QY	298 TLVQYLRPTAQYG-ADSPQWTFWSNPDGSGTGILKDDMQTV			339
Db	321 ALVDYL---ISIGVTDTP---YMSINANSSTGGLFRDDWNHV			357
RESULT 9				
Q9V052	PRELIMINARY;	PRT;	514 AA.	
AC	Q9V052	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)		
DE	Endoglucanase.			
GN	CEL8-LIKE OR PYRAB09400 OR PAB0632.			
OS	Pyrococcus abyssi.			
OC	Archaea; Euryarchaeota; Thermococci; Thermococcaceae;			
OC	Pyrococcus.			
OX	NCBI_TaxID=29292;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CE5 / Orsay;			
RA	"Pyrococcus abyssi genome sequence: insights into archaeal chromosome			
RT	structure and evolution."			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; AJ248285; CAB49854.1; ..			
DR	PIR; E75142; E75142.			
DR	HSSP; P54583; 1ECE.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001547; Glyco_Hydro_5.			
DR	Pfam; PF00150; cellulase; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Complete proteome.			
QY	SEQUENCE 514 AA; 59980 MW; CF80CAB1525337C5 CRC64;			
Query Match	34.3%; Score 682.5; DB 17; Length 514;			
Best Local Similarity	41.6%; Pred. No. 8.6e-45;			
Matches 148; Conservative	57; Mismatches 122; Indels 29; Gaps 10;			
QY	7 HTSGREILDANNV-----PVRIAGINWFGPFCNTVYVHGLWSRDRYRMLDQIKSLGYNVT			60
Db	30 YTAENGILFVQNVTTGKPKPLVHGVSWFGFELKDHVYVGLDKRKNKDKLKDVKRLGFNA			89
QY	61 IRLPYSDDLKPGTWNPN--SINFQNMQLQGLTSLQVMDKIVAYAGIQLRILDRHRP			118
Db	90 IRLPFCSESIRDPTRPSPERIN-YELNPDCLKLTSLKLEIKLIEVANSIGLYLLDYHRI			148
QY	119 DCSGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDW			178
Db	149 GCEETPELWYTNYSSEQYIKDWIFLAKRFGKYPNVIGADIKNEPHGEAGWGTGDER-DF			207
QY	179 RLAERAGNAVLSVNPNTLIFVEGVQ-----SYNGDSYV--WGNLQAGAGYVWL			227
Db	208 RLFAEKVGRKILKVAHPHLIFVEGVQYTHVPNIDIEIKKGGWTFWGNLQAGAGYVWL			267
QY	228 NVNRLVYSAHDYATSVGPOTWFSDDTFFNNMFGIKNWKGWYLFNQNIAPVWLGEETTL			287
Db	262 YGPDVYVQYFNSPDPFNNMAIWDKHFHFAKAGYA-MAIGFEGGKYGEGRDRAWQN			320
QY	298 TLVQYLRPTAQYG-ADSPQWTFWSNPDGSGTGILKDDMQTV			339
Db	321 ALVDYL---ISIGVTDTP---YMSINANSSTGGLFRDDWNHV			357

```
Db 262 YGPDVFVQSFNDNFNNPMDRHFQGFAGSH--ALLLGEFGKYGEGDARDKWQD 319
QY 298 TLVQYLRPTAQYAGDSFQWTFWNPDSGDTGGLKDDQWTV 339
Db 320 ALVKYLR---SKGIN--EGFYWNPNSGDTGGLLRDDWTSV 356

RESULT 6
Q58925 PRELIMINARY; PRT; 458 AA.
AC O58925;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 458AA long hypothetical endo-1,4-beta-glucanase.
GN PH1171.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3.
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamanoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takaiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb P.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL; AF000005; BAA30271.1; -.
DR PIR; E71059; E71059.
DR HGSP; P54583; LECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW Complete proteome.
SQ SEQUENCE 458 AA; 51930 MW; CAE48AD43A8EE654 CRC64;

Query Match 38.0%; Score 755; DB 17; Length 458;
Best Local Similarity 45.0%; Pred. No. 1.7e-50;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMLODKSLGYNTRILPYS 67
Db 54 TSGBE-----TFIHLFGVNWFGFETPNVHGLWKRNMEDMLLQKSLGFNAIRLPCT 107
QY 68 DILKPGTMPSINFYQNMNODLGSLTSLQVMDKIVAYAGIQRILDRHRPDCSGSALW 127
Db 108 ESKVPGTQPTIGD-YSKNPLRLGLSLQIMEKIIKAGDLGIFVLVDYHRICTHIEPLW 166
QY 128 YTSVSEATWISDLQALQRYKGNPTVGFDLNHEPHD-----PACWGGDPS 175
Db 167 YTEDFSEEDPTWIEVAKRFGKWNVIGADLNKNEPHTSVSPAATDGTGATWGNP 226
QY 176 IDWRLAERAGNAVLSVNPMLLI FVEGVQSYN-----GDSYVWNGNLQAGQYPVY 226
Db 227 TDNLAERIGKAILKAVPHLLIFVEGTFQPTNPKTSSYKWNYNMWNGLMAVKDYPV- 285
QY 227 LNVF-NRLVYSAHDYATSVGPQTFWSDPT-FNNMFGIINQWNGYLFNQNIAPVWLBEG 284
Db 286 -NLPRNKLVSHPVYGDVYVQYFPGAKFPNLPDIWYHFGYVKLELGYSVVIGEF 344
QY 285 TTL---QSTTDQWTLKTLVQYLRPTAQYAGDSFQWTFWNPDSGDTGGLKDDQWTV 339
Db 345 KGKYGCGDPEDVIQWKLVDWM--IENKFCDF--YWSNPNDSGDTGGLQDDWTI 397
```

RESULT 7

```
Q97KX6 PRELIMINARY; PRT; 482 AA.
ID Q97KX6;
AC Q97KX6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Possible non-processive endoglucanase family 5, secreted, Cella
DE homolog secreted, dockerin domain.
GN CAC0912.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe P., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007607; AAK78888.1; -.
DR PIR; E97012; E97012.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 482 AA; 54109 MW; 13E2E29197D59A5F CRC64;

Query Match 34.7%; Score 690; DB 16; Length 482;
Best Local Similarity 37.9%; Pred. No. 2.1e-45;
Matches 141; Conservative 71; Mismatches 126; Indels 34; Gaps 9;

QY 5 YWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMLODKSLGYNTRILP 64
Db 24 YLHSDGSKLLDDYGNQVRMTGIAWFGLETENYCPHGLWANRLDNILNIVADNGENTLRVP 83
QY 65 YSDDIL---KEGT--MPSINFYQNMNODLGSLTSLQVMDKIVAYAGIQRILDRHRPD 119
Db 84 LSVELVQWROGVYPTPDSINDY-ISPELKGQNSLQILDVVIAYSKVGVKVMQDMERIE 142
QY 120 CSGQSALWYTSVSEATWISDLQALQRYKGNPTVGFDLNHEPHDPAC-----WGCG 172
Db 143 SGGQTATWYTSKYTTDDDEYKQVYLDADRYKNDTVAADIFNEPHGKAYRAETSAKWN 202
QY 173 DPSIDWRLAERAGNAVLSVNPMLLI FVEGVQSYN-----NGDSY---WNGNLQ 219
Db 203 TDEDNRYEAEKVGKILIDINPKMLI VVEGVETIPKEGTAGSTNPDYVGGWGGNLRG 262
QY 220 AGQYFVVL-NVFNRLVYSAHDYATSVGPQTFWSDPTFNNMFGIINQWNGYLFNQNIAP 277
Db 263 VKDYFVDLAPYKNQVYSPHYDYGVSQDQWTFWNPDSGDTGGLKDDQWTV 322
QY 278 VWLGEFGTTLQSTTDQWTLKTLVQYLRPTAQYAGDSFQWTFWNPDSGDTGGLKDDQW 337
Db 323 LLIGEWGNMDDGKNQWMTDMAKLI-----SDKNMHTFWCLNANS GDTGGLILEYDFK 376
QY 338 TVDTVKDGYLAP 349
```

```

Db 165 SELWYTYEYEPESRWISDWRMLAKRYASDPTTVIGVDLHNEPHGAATWGTGAATTDWRAAE 224
Qy 184 RAGNAVLNVNPNLLIFVEGV-QSYNGDSYWMGNLQAGQYPPVNLVNPRLVYSADHYAT 242
Db 225 RGGNAVLNPKLLVLEGIDHQADGTGTWGGALDSAAATASVRLTVANRVVYSPPHYPS 284
Qy 243 SVGQOTWFSPTFPNNMGGIWNKNGYLFNQNIAPIWGLGFGTTLOSTDTQWTKTLVQY 302
Db 285 TIYGQWFSASNYPTNLPGIWDHAGWYLAKKDIAPVLVGFEGTKLETASDKQMLTLVGY 344
Qy 303 LRPTAQGADSFQWTFWSWNPDSGDTGGILKDWQTVTDVTKDGYLAPI 350
Db 345 LSST--GISS--SFAWNPDSGDTGIVKSDWVTEQAALDALAPI 386

RESULT 4
ID Q9EVR6 PRELIMINARY; PRT; 518 AA.
AC Q9EVR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extracellular endoglucanase (ENGXCA protein) precursor.
GN ENGXCA.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRL-B1459;
RX Schroeter K., Puchler A., Becker A.;
RA "engXCA major extracellular endoglucanase.";
RL Submitted (DSC-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ304415; CAC18529.1; -.
DR HSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL.
FT CHAIN 1 25 POTENTIAL.
FT FT 26 518 EXTRACELLULAR ENDOGLUCANASE (ENGXCA
FT PROTEIN).
SQ SEQUENCE 518 AA; 55568 MW; 9A0FF678E3BF712C CRC64;

Query Match 40.3%; Score 800.5; DB 2; Length 518;
Best Local Similarity 47.2%; Pred. No. 5.6e-54;
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;

Qy 5 YWHTSGREILDANNVPRIAGINWFGTCTNYVHGLWSRDYRSLDQIKSLGYNTRILP 64
Db 26 YSINNSQIVDDSKVQLKGVNFGTCTNYVHGLWLNWKNWDMVQMGGLGFAVRLP 85
Qy 65 YSDILKPGTTPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSGQS 124
Db 86 FCPATLRSDTWPASID-YSRNADLOGLTSLQILDKVIAEFNARGMYVLLDHTPDCAIS 144
Qy 125 ALWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPCWCGDPSIDWRLAER 184
Db 145 ELWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPCWCGDPSIDWRLAER 204
Qy 185 AGNAVLNVNPNLLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVNLVNP-NRLVYSAH 238
Db 205 GSAVLAVAPKWLIAVEGITDNPVCSTNG-GIFWGNLQPLACTP--LNIPANRLLIAPH 261
Qy 239 DYATSVGPOTWFSPTFPNNMGGIWNKNGYLFNQNIAPIWGLGFGTTL--QSTTDQTL 296
Db 262 VYGPVDFVQSYFNSNFPNNMGPALWERHFGQFAGTH--ALLGFEFGKYGEGDARDTKWQ 319
Qy 297 KTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDWQTV 339

```

```

Db 320 DALVXYLR---SKGIN--QGFWSWNPNSGDTGGILRDDWTSV 357

RESULT 5
ID Q8PPS3 PRELIMINARY; PRT; 474 AA.
AC Q8PPS3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase.
GN ENGXCA OR XAC0612.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardoso J., Chantengo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463 (2002).
DR EMBL; AB011689; AAM35501.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001519; Bac_celose-bind.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Complete proteome.
SQ SEQUENCE 474 AA; 51294 MW; 379B84D63CASA31B CRC64;

Query Match 40.2%; Score 799.5; DB 16; Length 474;
Best Local Similarity 46.5%; Pred. No. 5.9e-54;
Matches 159; Conservative 60; Mismatches 104; Indels 19; Gaps 9;

Qy 6 WHTSGREILDANNVPRIAGINWFGTCTNYVHGLWSRDYRSLDQIKSLGYNTRILP 65
Db 26 YSINNSQIVDDSKVQLKGVNFGTCTNYVHGLWLNWKNWDMVQMGGLGFAVRLP 85
Qy 66 SDDILKPGTTPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSGQSA 125
Db 86 FCPATLRSDTWPASID-YSRNADLOGLTSLQILDKVIAEFNARGMYVLLDHTPDCAISE 144
Qy 126 ALWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPCWCGDPSIDWRLAER 185
Db 145 ELWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPCWCGDPSIDWRLAER 204
Qy 186 AGNAVLNVNPNLLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVNLVNP-NRLVYSAH 239
Db 205 GSAVLAVAPKWLIAVEGITDNPVCSTNG-GIFWGNLQPLACTP--LNIPANRLLIAPH 261
Qy 240 YATSVGPOTWFSPTFPNNMGGIWNKNGYLFNQNIAPIWGLGFGTTL--QSTTDQTLK 297

```

```
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 156 SELWYTSQYPERWISDWKLAERYKNPTVIGADLHNEPHGQASWGTGVDSTWRLAAQ 215
QY 184 RAGNAVLSVNPVNLIFVEGVQ---SYNGDSYWMGNLQAGQYPVVNLVNPRLVYSADHY 240
Db 216 RAGNAVLSVNPVNLIFVEGVQDHNVKGNNSQYWMGNLQAGQYPVVNLVNPRLVYSADHY 275
QY 241 ATSVGPQWFSPTFPNNMGGIWNKNWGYLFNQTAPVWLGSEFG---TTLQSTTDCWTKT 298
Db 276 GFGVSSQWFSPTFPNNMGGIWNKNWGYLFNQTAPVWLGSEFG---TTLQSTTDCWTKT 335
QY 299 LVQYLRPTAQYGADSFOWTFWGNPDSDGTGILKDDWQTVTVKDGVLAPIKSIFFPV 358
Db 336 LVDYI-----GANNLYFYFWSLNFNSGDTGGLLDDWTTWNRPKQDML-----SRIMKPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
ID Q9K5C7
AC Q9K5C7
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OG Plasmid pCM1.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP STRAIN=NCPPB 382;
RC STRAIN=NCPPB 382;
RA Melzerus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-614 FROM N.A.
RC STRAIN=NCPPB 382;
RX MEDLINE=20331586; PubMed=10875331;
RA Jahr H., Dreier J., Meleuzus D., Bahro R., Eichenlaub R.J.;
RT "The endo-beta-1,4-glucanase Cella of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato."
RL Mol. Plant Microbe Interact. 13:703-714 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Jahr H.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR HSPB; X62582; CAA44467.2; -.
DR HSPB; P54583; IECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Iol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT SIGNAL 66 POTENTIAL.
FT CHAIN 67 >745 CELLULASE.
FT NON_TER 745
SQ SEQUENCE 745 AA; 77886 MW; 5845015EB8D4F27C CRC64;

Query Match 45.4%; Score 901.5; DB 2; Length 745;
Best Local Similarity 50.0%; Pred. No. 1.3e-61;
Matches 174; Conservative 51; Mismatches 112; Indels 11; Gaps 4;
```

```
QY 4 GYWHITSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRMLDQIKSLGYNTIRL 63
Db 70 GNLHTAGGKIVTAGAPYTIIRGIAWGMETSSCAPHGLDITITLGGMQHILKQMGFTTVRL 129
QY 64 PYSDDLKPGTWPNSINFYQNMQLGLTSLOVMDKIVAYAGQIGRLIILDRHPPDCSGQ 123
Db 130 PFSNQCL-----AASGVTGVSADPSLAGLTPLOVMDHVVASAKSAGLDVILDQHRPDSGGQ 185
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 186 SELWYTSQYPERWISDWKLAERYKNPTVIGADLHNEPHGQASWGTGVDSTWRLAAE 245
QY 184 RAGNAVLSVNPVNLIFVEGV-QSYNGDSYWMGNLQAGQYPVVNLVNPRLVYSADHYAT 242
Db 246 RGNNAVLAENLLVVEGIDHEADSGGTWGGALGLVGNAPVRLSVANRVVYSPHDYPS 305
QY 243 SVGPQWFSPTFPNNMGGIWNKNWGYLFNQTAPVWLGSEFGTTTOSTTDCWTKLTVQY 302
Db 306 TIYGSWFSASNPANLPGIWDHAGLYAKKDIPVLVGEFGTKFETTSKQWNLTVGY 365
QY 303 LRPTAQYGADSFOWTFWGNPDSDGTGILKDDWQTVTVKDGVLAPI 350
Db 366 LSSI---GISS---SPWAFNPNSGDTGGIVKSDWVTPEQAKLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
ID Q9AF65
AC Q9AF65;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cellulase Cella.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Laine M., Haapalainen M., Wahlroos T., Kankare K., Nissinen R.,
RA Kassuwi S., Mettler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain."
RL Physiol. Mol. Plant Pathol. 0:0-0 (2001).
DR EMBL; AY007311; AAK16222.1; -.
DR HSPB; P54583; IECE.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Iol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 727 AA; 75811 MW; 96DF9B664873985D CRC64;

Query Match 45.0%; Score 894.5; DB 2; Length 727;
Best Local Similarity 50.0%; Pred. No. 4.2e-61;
Matches 174; Conservative 48; Mismatches 115; Indels 11; Gaps 4;
```

```
QY 4 GYWHITSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRMLDQIKSLGYNTIRL 63
Db 49 GNLHTAGGKIVTAGAPYTIIRGIAWGMETSSCAPHGLDITITLGGMQHILKQMGFTTVRL 108
QY 64 PYSDDLKPGTWPNSINFYQNMQLGLTSLOVMDKIVAYAGQIGRLIILDRHPPDCSGQ 123
Db 109 PFSNQCL-----AASGVTGVDADPSLAGLTPLOVMDHVVASAKSAGLDVILDQHRPDSGGQ 164
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 ; Search time 36.3333 Seconds
(without alignments)
3108.867 Million cell updates/sec

Title: US-09-997-504A-10
Perfect score: 1987
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp archaea:*

2: sp bacteria:*

3: sp fungi:*

4: sp human:*

5: sp invertebrate:*

6: sp mammal:*

7: sp mhc:*

8: sp organelle:*

9: sp phage:*

10: sp plant:*

11: sp rodent:*

12: sp virus:*

13: sp vertebrate:*

14: sp unclassified:*

15: sp rvirus:*

16: sp bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1060	53.3	397	Q8RP23	Q8rp23 paenibacill
2	901.5	45.4	745	Q9K5C7	Q9k5c7 clavisbacter
3	894.5	45.0	727	Q9AF65	Q9af65 clavisbacter
4	800.5	40.3	518	Q9EVR6	Q9evr6 xanthomonas
5	799.5	40.2	474	Q8RPS3	Q8rps3 xanthomonas
6	755	38.0	458	17 C58925	C58925 pyrococcus
7	690	34.7	482	16 Q97FK6	Q97fk6 clostridium
8	687.5	34.6	592	16 Q9PF60	Q9pf60 xyella fas
9	682.5	34.3	514	17 Q9V052	Q9v052 pyrococcus
10	649.5	32.7	614	16 Q87AH4	Q87ah4 xyella fas
11	627.5	31.6	1000	2 C24820	C24820 thermophil
12	621	31.3	1426	2 Q9X3P6	Q9x3p6 caldicellul
13	620.5	31.2	660	2 Q9L3J2	Q9l3j2 clostridium
14	591.5	29.8	534	2 Q9EY00	Q9ey00 clostridium
15	582	29.3	616	2 Q7X2N2	Q7x2n2 thermomonos
16	513.5	25.8	425	2 Q93Q07	Q93q07 thermus cal

17	503	25.3	574	2 Q9S3V3	Q9s3v3 cellulomona
18	488	24.6	341	2 P96310	P96310 anaerocellu
19	403	20.3	630	2 Q8RJY7	Q8rjy7 stigmatella
20	328	16.5	565	16 Q9PF68	Q9pf68 xyella fas
21	315.5	15.9	569	16 Q87AG9	Q87ag9 xyella fas
22	276.5	13.9	590	16 Q8P513	Q8p513 xanthomonas
23	267	13.4	508	10 Q9LTM8	Q9ltm8 arabidopsis
24	258.5	13.0	522	10 Q9SAP6	Q9sap6 arabidopsis
25	221	11.1	526	10 Q9LFS2	Q9lfs2 arabidopsis
26	211.5	10.6	488	10 Q9LFE7	Q9lfe7 arabidopsis
27	196.5	9.9	555	10 Q7XUQ4	Q7xuq4 oryza sativ
28	190.5	9.6	354	16 Q8PFP2	Q8pfp2 xanthomonas
29	184.5	9.3	561	10 Q9LTM0	Q9ltm0 arabidopsis
30	173.5	8.7	437	3 Q99036	Q99036 trichoderma
31	170	8.6	550	10 Q7XUQ5	Q7xuq5 oryza sativ
32	169.5	8.5	694	2 Q9Z187	Q9z187 bacillus st
33	169	8.5	439	3 Q92401	Q92401 agaricus bi
34	168.5	8.5	377	3 Q00012	Q00012 aspergillus
35	168	8.5	439	3 Q9P893	Q9p893 agaricus bi
36	167	8.4	377	16 Q8PRD5	Q8prd5 xanthomonas
37	166.5	8.4	558	10 Q7XUQ6	Q7xuq6 oryza sativ
38	164.5	8.3	516	2 Q61855	Q61855 bacillus ci
39	160.5	8.1	357	16 Q8PRD3	Q8prd3 xanthomonas
40	160.5	8.1	722	16 Q97L56	Q97l56 clostridium
41	159.5	8.0	363	2 Q07652	Q07652 cellvibrio
42	157	7.9	364	2 Q9F0G8	Q9f0g8 rhizobium m
43	155.5	7.8	356	16 Q9PA12	Q9pai2 xyella fas
44	153.5	7.7	356	16 Q87X88	Q87x88 xyella fas
45	153.5	7.7	1449	10 Q8RU51	Q8ru51 oryza sativ

ALIGNMENTS

RESULT 1

Q8RP23 ID Q8RP23 PRELIMINARY; PRT; 397 AA.

AC Q8RP23; DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Endo-beta-1,4-glucanase

OS Paenibacillus sp. KCC8848P.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI_TaxID=109199;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=KCTC 8848P.

RA Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;

RT "Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-

expression with Endomyces fibuliger beta-glucosidase gene in

RT Saccharomyces cerevisiae."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF345984; AAL83749.1;

DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00150; cellulase; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

SQ SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;

Query Match 53.3%; Score 1060; DB 2; Length 397;

Best Local Similarity 55.0%; Pred No. 2, 4e-74;

Matches 198; Conservative 44; Mismatches 102; Indels 16; Gaps 5;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGPCTCNVYVHGLWSRDYRMLDQIKSLGNTRL 63

37 GYHTQGNKIVDETKEAFAFNWFLGTPNTYTLHGLWSRMDMLDQVKEGYNLRL 96

64 PYSDILKPGMPNSINFYQVNDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQ 123

97 PYSNQLFSDSSRADSIDYIK-NPDLVGLTPIQIMDKLIEKAGQGIQIILDRHRPDCSGQ 155

Db

This Page Blank (uspto)

402 QVWDNLSRGFTVITINSHDD-----WIKEDYNGNIEFEKXW---EQIAERFKNS 450
 153 TVVGFDLNPHDPACWCGDPSIDWLAERAGNAVLSVNPMLLIFVEGVQSYNGDSY 212
 451 ENLLFEIMNEPFG---NITDEQID--DMNSRIKIRKTNPRVII-----GGY 497
 213 WGNLQAGQVPPVNLVNSAHYATSVGPQWFSDPPTFPNNPGLWN----- 264
 498 -----NSYNTLVNI-----KIPDDPYLIGTFHYDPYETHKRWGTGQEDMDTV 543
 265 -----KMGYLFNQNIAPVWLGFEFTLQSTTDQWMLKTLVOYLRPTAQYQADSFQWT 317
 544 VRVDFVKWS---DRNIPVYFGEF--AVMAYADRTSRVKWYDFISDAAL--ERGFACS 596
 318 FWSNPPSGDTGGLKD-----DMQVDTVKGYLAPIKSSIFDP 357
 597 VW-----DNGVFGSLDNDMAIYNDRTRTFDT-----EILNALFNP 631

RESULT 15

GUN3_HUMIN STANDARD; PRT; 388 AA.
 AC Q12624; Q12620;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase 3 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 3)
 DE (Cellulase 3).
 GN CMC3.
 OS Humicola insolens.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 CX NCBI_TaxID=34413;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94247364; PubMed=8190078;
 RA Dalboege H., Hansen H.P.H.;
 RT "A novel method for efficient expression cloning of fungal enzyme
 genes.";
 RL Mol. Gen. Genet. 243:253-260(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO 9584;
 RA Takashima S., Nakamura A., Masaki H., Uozumi T.;
 RT "Cloning, sequencing and expression of the thermostable cellulase
 gene of Humicola grisea.";
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -! BIOTECHNOLOGY: Used as a detergent cellulase. Sold under the name
 Cellzyme by Novozymes. This special enzyme has three effects:
 CC colour brightening, softening and removal of particulate soil. The
 CC overall effect is that it helps to preserve the nice appearance of
 CC new fabric and restores old fabric so that it looks new again.
 CC -! SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 hydrolases).
 CC -! SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X76046; CAA53631.1; -;
 CC EMBL; D84470; BAA12676.1; -;
 CC PIR; S43920; S43920.
 CC HSP; P00725; IAZ6.
 CC InterPro; IPR00254; CBD_fungal.
 CC InterPro; IPR001547; Glyco_hydro_5.
 CC Pfam; PF00734; CBM_1; 1.
 CC Pfam; PF00150; cellulase; 1.

DR ProDom; PD001821; CBD_fungal; 1.
 DR SMART; SM00236; fCBD_1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR PROSITE; PS00569; GLYCOSYL_HYDROL_F5; FALSE_NEG.
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 388
 FT DOMAIN 18 52
 FT DOMAIN 53 91
 FT DOMAIN 92 388
 FT DISULFID 24 41
 FT DISULFID 35 51
 FT CARBOHYD 92 92
 FT CARBOHYD 155 155
 FT CARBOHYD 259 259
 FT ACT_SITE 215 215
 FT ACT_SITE 322 322
 FT CONFLICT 8 8
 FT CONFLICT 340 340
 FT SEQUENCE 388 AA; 42563 MW; C7CF349DACC10690 CRC64;
 Query Match 6.9%; Score 136.5; DB 1; Length 388;
 Best Local Similarity 25.5%; Pred. No. 0.0022;
 Matches 84; Conservative 38; Mismatches 134; Indels 73; Gaps 21;
 QY 28 WFGP-ETC-----NYVHGLWSRDY-----RSMLDQIKSLGYNTIRLPYSDDILKPGTMP 76
 DB 88 WFGINQSCAEFGKGEY--PGLWGXHFTPPSTSSIQTHINDGFNMFVAFSMERL----AP 141
 QY 77 NSINEFYQNNQDLQGLTSLQVMDKIVAVAGQGLRIILDRHRPDCSGQSALVYTSVSE-A 135
 DB 142 NOLNAAFDANYLRNL-----ETVNFITGKGVKAMLDPH-----NFGRYEIRIIDKA 189
 QY 136 TWISDLQALQRYKGNPTVVGFDLHNEPHD---PACWCGDPSIDWLAERAGNAVLSV 192
 DB 190 AFASFTTKLATHFASNPLVW-FDTNNEYHDMQQLVFDLNOQAID---AIRAAGAT--- 241
 QY 193 NPNLLIFVEGVQSYNGDSY--WGNLQAGQYPPVNLVNP-NRLVYSAHDYATSVGPQTW 249
 DB 242 --SQYIMVEG-NSWTGAWTNVNTNNL-----AALRDPENKLVYQHQVLDSDSGSGT- 290
 QY 250 FSDPTFPNNMPG-----IWNKNWGLFNQNIAPVWLGFEFTLQSTTDQWMLKTLVQYLRP 305
 DB 291 --STACVSTQVGLQVRVIGATNW---LRQNGKVGILLGEFAGGANSVCQA-IEGMLTHLQE 344
 QY 306 TAYQADSFQWT-----FWSNPPDSG 326
 DB 345 NSDVWTGALWAGGPPWGDYIYSEPPSG 373

Search completed: August 2, 2004, 16:34:55
 Job time : 16.3333 secs

```

CC      WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC      MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC      -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC      hydrolases).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; D90341; BA014354.1; --
CC      HSSP; P17901; LEDG; CBM 11.
CC      InterPro; IPR005087; CBM 11.
CC      InterPro; IPR002105; Dockerin_1.
CC      InterPro; IPR002048; EF-hand.
CC      InterPro; IPR001547; Glyco_hydro_5.
CC      Pfam; PF03425; CBM 11; 1.
CC      Pfam; PF00150; cellulase; 1.
CC      Pfam; PF00404; Dockerin_1; 2.
CC      PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC      PROSITE; PS00448; CLOS CELLULOSE RPT; 1.
CC      PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC      Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC      SIGNAL 1 24 POTENTIAL.
CC      CHAIN 25 584 ENDOGLUCANASE D.
CC      DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
CC      DOMAIN 329 353 PRO/THR-RICH (LINKER).
CC      DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
CC      DOMAIN 530 584 2 X 24 AA APPROXIMATE REPEATS.
CC      REPEAT 530 552 1.
CC      REPEAT 562 584 2.
CC      ACT_SITE 159 159 PROTON DONOR (BY SIMILARITY).
CC      ACT_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).
CC      SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;

Query Match 7.1%; Score 140.5; DB 1; Length 584;
Best Local Similarity 21.0%; Pred. No. 0.0018;
Matches 71; Conservative 52; Mismatches 130; Indels 85; Gaps 18;

QY 21 VRIAGINWGFETCNVTVHGLMSRDYRS-MLDQIKSLGNTYRLPYSDILKPGTWPNISI 79
DB 32 VKRMGIGMNLGNTFDAPTEGSKAAQAYYDFDFKQAGFKHVRIPIRWDQHTLANSPTV 91
QY 80 NPYQMNQDLQGLTSLQVMPKIVAYAGQIGLRIILDRHRPDCSGQSAW-----YTSSVS-- 133
DB 92 DSNFLNR-----IETVIDMSLSRGFTVTVINSHD-----TWLMDNYSQNICRF 134
QY 134 EATWISDQLAQRKYGNFTVGVDFLHNEPHDPACWCGCDPSIDWELAAERAGNAVLSVN 193
DB 135 EKIW-----EQIAQRFKGKSENLFVEILNEPHG-----NITDSQIN--DMNKRILNIIRKTN 184
QY 194 PNLLIEFVEGVQSYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDYATSVGPQTWFSDDP 253
DB 185 PFRNVII-----GAGYWNYSNLSQLIP---NDPN-LIATFYV-----DP 222
QY 254 -TFPNMNPGLWN-----KNGYLFNQNIAPIVWLGFGFTLQTTQTTQWLK 297
DB 223 YSFTHWQQTGTGKTNDMDAIAVFNHVKWS---DKNNIPVYLGEYV--VMGHSRTSAV 277
QY 298 TLVQYLRPTA-QYGADSFQWTFWSNPNDSGDTGGILKD 334
DB 278 KWFDFVSDQAIHGFGCGAW-----DNGVFGSVND 308

RESULT 14
ID GUNH CLOTM STANDARD; PRT; 900 AA.
AC P16218;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase)
DE (Cellulase H).
GN CELH.
OS Clostridium thermoCELLUM.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCIB 10682;
RX MEDLINE=90323606; PubMed=2197182;
RA Yacquee E., Beguin P., Aubert J.-P.;
RT "Nucleotide sequence and deletion analysis of the cellulase-encoding
RT gene celH of Clostridium thermoCELLUM.";
RL Gene 89:61-67(1990).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 26 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31903; AAA23225.1; --
CC PIR; JH0157; JH0157.
CC HSSP; P07985; 1CEC.
CC InterPro; IPR005087; CBM 11.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR002048; EF-hand_5.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF03425; CBM 11; 1.
CC Pfam; PF00150; cellulase; 1.
CC Pfam; PF00404; Dockerin_1; 2.
CC PROSITE; PS00018; EF HAND; UNKNOWN 1.
CC PROSITE; PS00448; CLOS CELLULOSE RPT; 2.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC SIGNAL 1 44
CC CHAIN 45 900 ENDOGLUCANASE H.
CC DOMAIN 45 630 CATALYTIC (BY SIMILARITY).
CC DOMAIN 631 654 PRO/THR-RICH (LINKER).
CC DOMAIN 655 900 CELLULOSE-BINDING (BY SIMILARITY).
CC ACT_SITE 460 460 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 565 565 NUCLEOPHILE (BY SIMILARITY).
CC DOMAIN 833 895 2 X 24 AA APPROXIMATE REPEATS.
CC REPEAT 833 856 1.
CC REPEAT 872 895 2.
CC SEQUENCE 900 AA; 102415 MW; 973AFB1954FC246B CRC64;

Query Match 6.9%; Score 137; DB 1; Length 900;
Best Local Similarity 20.2%; Pred. No. 0.0057;
Matches 70; Conservative 51; Mismatches 131; Indels 94; Gaps 18;

QY 40 GLWSDR-YRSMLDQIKSLGNTYRLPYSDILKPGTNPNISINFYQMNQDLQGLTSLQVMD 98
DB 352 GSWKSAMEYVDFDKAAGYKQVRIPIVRWDNHTWRTYPTID-----KAFLDRIE 401
QY 99 KIVAYAGQIGLRIILDRHRPDCSGQSAW-----YTSSVS--EATWISDQLAQRKYGNP 152
DB 99 KIVAYAGQIGLRIILDRHRPDCSGQSAW-----YTSSVS--EATWISDQLAQRKYGNP 152

```

Query Match 8.1%; Score 161.5; DB 1; Length 814;
Best Local Similarity 21.8%; Pred. No. 6.3e-05;
Matches 84; Conservative 59; Mismatches 132; Indels 111; Gaps 20;
QY 6 WHTSGREILDANNVPVRIAG-----INWFGFETCNVYVHGLWS--RDYRSM 49
DB 32 FQVSGTKLLDASGNELVMGRDISALDLVKEIKGNLGNLTLDAPTETAWGNPRTRAM 91
QY 50 LDOIKSLGNTYITLPSYDILKPGTMPSINFYQMNQDLQGLTSLOVMKIVAYAGQIGL 109
DB 92 IEKVRMGFNAVPVPTWD--THIGPAPD---YKIDE-----AWLNRVSEVNVYLDQGM 141
QY 110 RIILDRHRPDCSGSALWYTSVSEATWISDLOA-----LAQYKGNPTV 154
DB 142 YALINLHH-----DNWILPTVYANRQSRKEKLVKWEQIATRFKDYDDH 185
QY 155 VGFDLNHPHPACWCGDPSIDWRLAERAGNAVLNVNPNLLIFVGVQSYNGDSYWWG 214
DB 186 LLEFETWNEPRE-----VGSP-MEWMGTYENRDNVIRFN---LAVVNTIRAS-----G 229
QY 215 GN-----LOGAGQYFVILN---VFN---RLVYSAHDY-----ATSV-GPQTWFS- 252
DB 230 GNDKRPILVPTNAATGLDVALNDLVPNNDSRVIVSIHAYSPYFFAMDVNGTSYWSY 289
QY 253 --PTFPNNMGIWNKNGYLFNQNIAFPVWLGFEFGTTLQSTTDQTLKTLV-----QYLREP 306
DB 290 DKASLSLSELDIYNR-----FVNGRAVLIIGF-----TIDKNLSRVAHAHYAREA 339
QY 307 AQYGADSFQWTFWNPDSGDTGIL 332
DB 340 VSRGIAVFWMDNGYVNPDAETVALL 365

RESULT 12
GUNI_RUMAL STANDARD; PRT; 406 AA.
AC P16216;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EG-I).
GN EG I.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.
RC STRAIN=F-40;
RX MEDLINE=90078126; PubMed=2687251;
RA Ohmura K., Kajino T., Kato A., Shimizu S.;
RT "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene.";
RL J. Bacteriol. 171:6771-6775(1989).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30928; AAA26469.1; .
DR PIR; A43722; A43722.
DR HSRF; P17501; 1EDG.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR

Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL 1 43
FT CHAIN 44 406 ENDOGLUCANASE I.
FT ACT_SITE 210 210 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 330 330 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45390 MW; 2E0172437B14FEA8 CRC64;
Query Match 7.4%; Score 147.5; DB 1; Length 406;
Best Local Similarity 22.9%; Pred. No. 0.00033;
Matches 88; Conservative 49; Mismatches 152; Indels 95; Gaps 18;
QY 15 DANNVPVRIAGIN-----WFGFETCNVYVHGLWSR-----DYR 47
DB 49 ETENVPVSQTHNDTMTVTSAKDLVAKMTNGWNLGNTMDATAQGLGSEVSWLPKVTNK 108
QY 48 SMLDOIKSLGNTYITLPSYDILKPGTMPSINFYQMNQDLQGLTSLOVMKIVAYAGQI 107
DB 109 YNIDMLPEAGNVLRIPVS-----WGNHIDDKYTSDFANMDRVQ---EIVNYGIDN 157
QY 108 GLRIILDRHRPDCSGSALWYTSVSE-----ATWISDLOALQRYKGNPTVWVGF 158
DB 158 GLYVILNTHHEE-----WTMPKPEKDGIEIEIKAWAQ-----IADRFKGYDEHLIFE 206
QY 159 LHNPHDPACWCGDPSIDWRLAERA-----GNVLNVNPNLLIFVGVQSYNGDS 210
DB 207 GLNEP-----RLRGEGAEWTGTSAREIINEYEKAFVETVRASGNGNDR 251
QY 211 YWVGNGNLQAGQYF---VVLNVF---NRLVYSAHDYAT---SVGPQTWFSPTFPNN--MPGI 262
DB 252 CLMITGYAASAYNNLSAIELPEDSKLIISVHAYLPVSPALDTKGTDKYDEPDTAIPEL 311
QY 263 WNKNGWYLFNQNIAFPVWLGFEFGTTLQSTTDQTLKTLVQLRPTAQYGADSFQWTFW 322
DB 312 F-EHLNELFISKGIPVIVGFEFGTWNKENTEDR-VKCLEDYLAATAAKYDIPCVWMDNYA-R 368
323 PDSGDTGGILKD---DMQTVDTVK 343
369 IONGENFGIMNRADLEWYFPDLIE 392
GUND_CLOCE STANDARD; PRT; 584 AA.
ID GUND_CLOCE
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)
DE (Cellulase D) (EGCCD).
GN CELCCD.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RX MEDLINE=92009193; PubMed=1916275;
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celCCD, of Clostridium cellulolyticum.";
RL Gene 104:33-38(1991).
CC -!- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellulohydrolases that cut the disaccharide cellobiose
CC from the nonreducing end of the cellulose polymer chain; (3)
CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other
CC short cello-oligosaccharides to glucose.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PATHWAY: Cellulose degradation.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS

RP SEQUENCE FROM N.A.
RC STRAIN=AW;
RX MEDLINE=92138626; PubMed=1735723;
RA Huang J., Schell M.A.;
RT "Role of the two-component leader sequence and mature amino acid
sequences in extracellular export of endoglucanase EGL from
Pseudomonas solanacearum";
RL J. Bacteriol. 174:1314-1323 (1992).
RN [2]
SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=89291722; PubMed=2738021;
RX Huang J., Sukordhaman M., Schell M.A.;
RA "Excretion of the egl gene product of Pseudomonas solanacearum";
RL J. Bacteriol. 171:3767-3774 (1989).
RN [3]
PROCESSING.
RP MEDLINE=90307678; PubMed=2195024;
RX Huang J., Schell M.A.;
RA "Evidence that extracellular export of the endoglucanase encoded by
egl of Pseudomonas solanacearum occurs by a two-step process
involving a lipoprotein intermediate";
RL J. Biol. Chem. 265:11628-11632 (1990).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Probable).
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M84922; AAA61980.1; -
CC PIR; A42649; A42649.
CC InterPro; IPR001547; Glyco hydro 5.
CC InterPro; IPR000437; Prok Lipoprot_s.
CC Pfam; PF00150; cellulase_1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
KW Lipoprotein; Signal; Palmitate.
FT SIGNAL 1 19
FT PROPEP 20 45
FT CHAIN 46 426
FT LIPID 20 20
FT LIPID 20 20
FT ACT SITE 249 249
FT ACT SITE 361 361
FT SEQUENCE 426 AA; 45578 MW; 51E13AD442CF4A8 CRC64;
Query Match 8.8%; Score 175; DB 1; Length 426;
Best Local Similarity 23.7%; Pred.No. 2.6e-06;
Matches 84; Conservative 55; Mismatches 135; Indels 80; Gaps 19;
QY 25 GINWFGFTCTNYVHGLWSDY-----RSMLDQIKSLGYNTIRLPYSDDIKPGTWPNSIN 80
DB 121 GVSLAGAEFGESLPCTGYSNYTPSADSVTYVKNKGMVLPLFRWERLPQ----- 172
QY 81 FVQMNC--DLQGLTSLOWDKIVAYAGQIGLRIILDRHPPDCSGOSALWY-----TSSVS 133
DB 173 --TLNFFDANELUSRTGTFVNAVATGQ---TVLLDPH-----NYARYGNVIGSSAVP 221
QY 134 EATWISDQALQARYKNGTFTVGFDFLHNEPHFPACWCGDPSIDWRLAERAGNAVLSVN 193
DB 222 NSAYADFWRRLATQFKSNRVI-LGLMNEPNSN-----PTEQMLSGANAEALAIRSAN 273
QY 194 PNLLIFVEGVQSVNGSDYVWGNLQAGQVPVY--LNVP-NRLIVSAHDY----- 240

DB 274 ASNVFVPG-NAWTG-AHSWNQWYGTPTNGTVMKGINDPGHNLVFEVHQVLDGSSGQSA 331
QY 241 ---ATSVGPOTWFSDDPTFPNNMPGINKKNGWYLFNQNIAPIWLGEGTTLQSTDTQTLWK 297
DB 332 NCVSATIGAQR-LQDFT-----TWLRSNGY-----RGFLGEGFAASNDTCNQA-VS 375
QY 298 TLVOYLRPTAQYGADSFQWTFWKNPDSGDTGGILK--DDWQTVDTVKDGYLAP 349
DB 376 NMLTFVKNNAADV-----WTGNAWAGGPWVGWGYMYSIEPSNGVDKPFQMSVLAP 423
RESULT 11
ID GUNE CLOTM STANDARD; PRT; 814 AA.
AC P10477;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase E precursor (EC 3.2.1.4) (EGE) (Endo-1,4-beta-glucanase)
GN (Cellulase E).
CN CELE.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RX MEDLINE=89137992; PubMed=3066698;
RA Hall J., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "Conserved reiterated domains in Clostridium thermocellum
endoglucanases are not essential for catalytic activity";
RL Gene 69:29-38 (1988).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M22759; AAA23224.1; -
CC PIR; JTO347; CZCLEM.
CC HSSP; P17901; 1EDG.
CC InterPro; IPR002105; Dockerin 1.
CC InterPro; IPR001547; Glyco hydro 5.
CC InterPro; IPR001087; Lipase_GDSL.
CC Pfam; PF00150; cellulase_1.
CC Pfam; PF00404; Dockerin_1; 2.
CC Pfam; PF00657; Lipase_GDSL; 1.
CC PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 34
FT CHAIN 35 814
FT ACT SITE 193 193
FT ACT SITE 316 316
FT DOMAIN 415 474
FT REPEAT 415 438
FT REPEAT 451 474
FT SEQUENCE 814 AA; 90244 MW; C6FA24B8D1523632 CRC64;

DOMAIN 281 748 CATALYTIC.
 FT DISULFID 39 133 BY SIMILARITY.
 FT ACT SITE 503 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 653 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;
 Query Match 14.7%; Score 291.5; DB 1; Length 748;
 Best Local Similarity 23.4%; Pred. No. 4.5e-15;
 Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;
 QY 7 HTSGREITLANNVPVRIAGNWFGEFCN-----YVHGLW-----S 43
 Db 290 NTKGNLTGQQLLPARG--NWFLEGRHEPSNDADNPSGAPMELVAGNMMWNNSSQSG 347
 QY 44 RDRSRMLDQIKSLGYNTIRLPYSDIL---KPTWPNNSINFQNMQDLOGLTSLQVMDK 99
 Db 348 RTIQQTWELKQOQITWLRPIAPQTLANDPQGRSPNLKN---HQSIQSNARQALD 403
 QY 100 IVVAGQIGLRIILDR-----RPDCSGOSALWY-----TSSVS 133
 Db 404 FIKLADQNDIQIFIDHSCSNYVGRAGRLDARPPYVDNRVGYDTTREYSCSAINNS 463
 QY 134 EAT-----WISDLQALA--CRYKGNPTVVGFDLHNEPDPACWGCDDPSIDWRLAA 182
 Db 464 SVTRIHAQKQKMLNRLREIAGLSAKLVSNLIGIDVFNEFYD-YTWA-----EWKGMV 516
 QY 183 ERAGNVLNPNLLIFVEGVQSYNGDSY-----WWGGLNLCAGQY 223
 Db 517 EEAQYAINNEVNNMLIIVEGI-SANANTDGTPTSVVPVPHGSTDLNPNNGENLYEAGAN 575
 QY 224 PVLNVP-NRLVYSAHDYATSVGPOTWFSPT-----FPNNMPG 261
 Db 576 PP-NIPKRLRSPHYGYSVFQRFMDPAQTECAGLEGDEAAQARCRIVINPTVLEQ 633
 QY 262 IWNKNGYLFNONTAPVWLGEFG-----TTLOSTTDQWTKTLVQYL 303
 Db 634 GWSEHFGYRELGYG-LIIEFGGNDMPGAKSSQADRNASHITTVDDQWQQAASYF 692
 QY 304 RPTAQYAGDSFQWTFWPNPDSGDTGILKDDQCTVDTVTKDGLYLAIPKISIFDPV 358
 Db 693 K---RKGINA---CYWSNPEASDTMG-----YLP-----WDPV 722

RESULT 9
 GUN1_RALSO STANDARD; PRT; 424 AA.
 ID GUN1_RALSO STANDARD; PRT; 424 AA.
 AC P58599;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase)
 GN EGL OR R5P0162 OR R505516.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin D., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

(Probable).
 -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 hydrolases).
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL curation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 EMBL; AL646076; CAD17313.1; ALT INIT.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR InterPro; IPR000437; Prok_Lipopr_t_S.
 DR Pfam; PF0150; cellulase; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
 KW Lipoprotein; Signal; Plasmid; Complete proteome; Palmitate.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 43 BY SIMILARITY.
 FT CHAIN 44 424 ENDOGLUCANASE.
 FT LIPID 20 20 S-diacylglycerol cysteine (By
 similarity).
 FT LIPID 20 20 N-palmitoyl cysteine (By similarity).
 FT ACT_SITE 247 247 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 359 359 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;
 Query Match 9.4%; Score 187; DB 1; Length 424;
 Best Local Similarity 25.0%; Pred. No. 3e-07;
 Matches 85; Conservative 51; Mismatches 122; Indels 82; Gaps 19;
 QY 35 NYVHGLRSRYSRMLDQIKSLGYNTIRLPYSDILKPTWPNNSINFQNMQ--DLOGLT 92
 Db 139 NYIYPSASATY-----YKNGMNLVLPFRWELQP-----TLNQALDANELS 182
 QY 93 SLQVMDKIVAYAGQIGLRIILDRHRRPDCSGOSALWY-----TSSVSEATWISDLQALAQR 147
 Db 183 RLTFGVNAVTAAGQ--TVLLDPH-----NVARYYGVNIGSSAVPNSAYADFRRVATQ 233
 QY 148 YKGNPTVVGFDLHNEPDPACWGCDDPSIDWRLAAERAGNVLNPNLLIFVEGVQSYN 207
 Db 234 FKGNAVY-FGLMNEPNSM-----PTEQWLSGANALAAIRSANASNVVFPVPG-NAWT 284
 QY 208 GDSYVWGGNLCAGQYVYV--LNVPNR-LVYSAHDY-----ATSVGPQTWFS 251
 Db 285 G-AHSMNQNWYGTPTNGVTWKGINDPGRNLVFEVHQYLDGSSGOSASCVSATIGAERL-- 341
 QY 252 DPTFNNMPGIVNKGWYLFNONTAPVWLGEFGITLQSTTDQWTKTLVQYLRTPAQYGA 311
 Db 342 -QDFTN-----WLRNGY-----RGLGFBGAASNDTCNQA-VANMLTFVKNNAVY-- 385
 QY 312 DSFQWTFWSPNPDGDTGILK--DDWQTVDTVTKDGYLAP 349
 Db 386 ----WIGAWWAGGFWGMYVSIPEPSNGVDKPKQMSVLAP 421
 RESULT 10
 GUN2_RALSO STANDARD; PRT; 426 AA.
 ID GUN2_RALSO STANDARD; PRT; 426 AA.
 AC P17974;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN EGL.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (EGB) (Endo-1,4-beta-glucanase)
 DE (Cellulase B).
 GN CELB.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=86148508; PubMed=34531202;
 RA Grepinet O., Beguin P.,
 RT "Sequence of the cellulase gene of Clostridium thermocellum coding
 RL Nucleic Acids Res. 14:1791-1799(1986).
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03592; CAA27266.1; -.
 DR PIR; A23512; CZCLBM.
 DR HSSP; P54583; 1ECE.
 DR InterPro; IPR002105; Dockerin 1.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00150; cellulase; 1.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
 DR PROSITE; PS00448; CLOS CELLULOSE RPT; 2.
 DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 KM SIGNAL 1 27 OR 31
 FT CHAIN 28 563 ENDOGLUCANASE B.
 FT ACT_SITE 204 204 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 502 557 2 X 24 AA APPROXIMATE REPEATS.
 FT REPEAT 502 526 1.
 FT REPEAT 534 557 2.
 SQ SEQUENCE 563 AA; 63929 MW; 866FE55704A1D84B CRC64;
 Query Match 26.3%; Score 522; DB 1; Length 563;
 Best Local Similarity 30.7%; Pred. No. 3.9e-33;
 Matches 129; Conservative 58; Mismatches 159; Indels 74; Gaps 13;
 QY 1 AGGYY-----W-HTSGREILDANNVPRVITAGINWFGFCNVCVHGLWSRDYRSM 50
 Db 27 AEGSYADLAPDDDLHVEGTNIVDKYGNKWTGAWGFCNCRERMILDSVHSDIADI 86
 QY 51 DOIKSLGYNTIRLPYSDIIL---KPGTMPNSINFYQNOODLGTSLOWNDKIVAVAGOI 107
 Db 87 ELVADKGINVVRMPFIATDILLYAWSQGIYPSTDTSYNNPALAGLNGVELFNFLNFKRV 146
 QY 108 GLAIILDRHPPDCGQS---ALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH 164
 Db 147 GIKVILDVHSPETNDQGHVPLWNTITITIEIFKKAWVVAERYKNDTIIGFDLKNNEH 206
 QY 165 -----DPACWGCGDFSDIRLAAERAGNAVLSVNPNNLLIFVEGVQSY----- 206

Db 207 TNTGTMKIKQAISWDSNHPNNWKEVAETALAILVEHPNVLI FVEGVEMYPKOGIWDD 266
 QY 207 -----NGDSY--WMGNGLQAGAGQYPVVL--NVPNRLVYSADHYATISVGPOTWFSDP 253
 Db 267 ETEDTSPWTCGNDYGNWNGNLRGVKDYPIINLGKYQSOLVSPHPDYGPIVVEQDFKGD 326
 QY 254 TFPNN-----MPGIWKNWGYLFENONIAPIWVLGCEFGTTLQSTTDQTWLKTLVOYLRP 305
 Db 327 FITANDEQAKRIYEQCRDNWAYIYEBGISPILGEGWGMTEG--GHPILDLNKLKLR 384
 QY 306 TAOYGADS---FQWTFWSMNPDSGDTGGILKDP-----WQTVTVKDGVLAIPI 350
 Db 385 MRDFILENKYKLHHTFWCINIDSADTGGILFTRDEGTPFGGRDLKWN--DNKYDNLVYFV 442
 RESULT 8
 GUNC_PSEFL ID_GUNC_PSEFL STANDARD; PRT; 748 AA.
 AC P27033;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase) (EGC) (Cellobextrinase C).
 GN CELC.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OC NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.
 RC STRAIN=Sp; Cellulosa;
 RX MEDLINE=92061996; PubMed=1953673;
 RA Ferreira L.M.A., Hazlewood G.P., Barker P.J., Gilbert H.J.;
 RT "The cellobextrinase from Pseudomonas fluorescens subsp. cellulosa
 RL Biochem. J. 279:793-799(1991).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
 CC domain.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X61299; CAA43597.1; -.
 DR PIR; S19652; S19652.
 DR HSSP; P07986; 1EXG.
 DR InterPro; IPR001919; Bac celose-bind.
 DR InterPro; IPR009031; CBD.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR002883; Dockerin CBD_5.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF02013; CBM 10; 1.
 DR Pfam; PF00553; CBM_2; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR SMART; SM00637; CBD II; 1.
 DR PROSITE; PS00561; CBD BACTERIAL; 1.
 DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 38
 FT CHAIN 39 748 ENDOGLUCANASE C.
 FT DOMAIN 39 136 CELLULOSE-BINDING.
 FT DOMAIN 137 179 SER-RICH (LINKER).
 FT DOMAIN 228 280 SER-RICH (LINKER).

DR Pfam: PF00404; Dockerin 1; 2.
 DR PROSITE; PS00018; EF HAND; UNKNOWN 2.
 DR PROSITE; PS00448; GLOS CELLULOSOME RPT; 2.
 DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 566
 FT ACT_SITE 226 226
 FT ACT_SITE 381 381
 FT DOMAIN 503 549
 FT REPEAT 503 526
 FT REPEAT 536 549
 SQ SEQUENCE 566 AA; 63199 MW; 2CC9DE1AD87C3178 CRC64;
 Query Match 30.2%; Score 600.5; DB 1; Length 566;
 Best Local Similarity 34.2%; Pred. No. 3.2e-39;
 Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;
 Qy 6 W-HTSREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLP 64
 Db 44 WLHCKGNKIYDMYGNVWLTGANWFGNCFHGAW-YDVKTILTSIADRGINLLRIP 102
 Qy 65 YSDIIL-----KPG-----TMSINIFYQMNDLQ-----LTSLOWMDKIVAYAGQI 107
 Db 103 ISTEILYSMMIGKPNVSVSTASNNPPHYVNPDPYDPTDDVKNSEIFDIIMGYCKEL 162
 Qy 108 GLRIILDRHRPDC--SQGS-ALWY---TSS---VSEATWISDLQALQRYKGNPTVVGF 158
 Db 163 GIKWIDIHSDPANNSSGHNYELWYKSTCGVVTWKWIDTLWLADKYKNDITIAED 222
 Qy 159 LHNEPHDP-----ACWCGGDSIDWRLLAERAGNAVLSVNPILLIFVEGVQSY- 206
 Db 223 LKNEPHGRGYTAEPVPLKAKWNDSTDENNNKAAETCAKAILVNPVKVLIVTEGVEQP 282
 Qy 207 -----NGD-----SYWNGNLOAGQYPPVVLNPN-RLVYSADHYATSV 244
 Db 283 KTEGYTYDTPDWGATGDASPPYSANWGNLRGVKDYDIDGLPLNSQIVYSHDVGPSV 342
 Qy 245 GPQTFW-SDTPFPNNMGINWKNWGYLFNQNIAPVWLGEGFTTLQSTDTOTWKLTLVQYL 303
 Db 343 YAPWPEKEDFTMTQLDDYVDTWAYTHDQIAPILIGEMGHGMDGKGNKWNLLRDYI 402
 Qy 304 RPTAQYADSFQWTFWSPDSDGTGGLKDDQVTDVTKDGLAP 349
 Db 403 -----VQNRHHTFWCINPNSGDTGGLGNDWSTWDEAKYALLKP 442

RESULT 6

GUND CELFI
 ID_GUND_CELFI STANDARD; PRT; 747 AA.
 AC PS0400;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN CEND
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococineae; Cellulomonadaceae; Cellulomonas.
 OC NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93205933; PubMed=8458833;
 RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 RT D (cend), a family A beta-1,4-glucanase."
 RL J. Bacteriol. 175:1910-1918 (1993).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -!- PATHWAY: Cellulose degradation.
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
 CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

CC domain.
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
 CC hydrolases).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L02544; AAA23089.1; -.
 DR HSSP; P07986; 1EXG.
 DR InterPro; IPR001919; Bac celose-bind.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR008967; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF00553; CBM 2; 1.
 DR Pfam; PF00041; fn3; 2.
 DR SMART; SM00637; CBD II; 1.
 DR SMART; SM00060; FN3_2.
 DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 39
 FT CHAIN 40 747
 FT DOMAIN 448 542
 FT DOMAIN 546 639
 FT ACT_SITE 208 208
 FT ACT_SITE 349 349
 SQ SEQUENCE 747 AA; 78936 MW; BDL5473C9D8B42BD CRC64;
 Query Match 27.7%; Score 549.5; DB 1; Length 747;
 Best Local Similarity 35.0%; Pred. No. 4.1e-35;
 Matches 132; Conservative 58; Mismatches 142; Indels 45; Gaps 15;
 Qy 1 AGGQVW-HTSREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYN 59
 Db 40 ATGDDWLHVEGNTIYDSTGKEALLSGVNWFGFNASERVFHLWSGNITQITQQMAQRGIN 99
 Qy 60 TRLPYSDIIL---KPGTW---PNSINIFYQMNDLQGLTSLQWMDKIVAYAGQIGRIILD 114
 Db 100 VYRVFVSTQLLEWKAFTLKP-NVY-ANPELEKNSLQIFEWLTLCKYGIKVPFLD 157
 Qy 115 RH--RPDCSGQ-SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPH-----D 165
 Db 158 VHSAEADNSGHVYNNWKGDIITTEDVYEGWEWAATRWKDDDTIVGADIKNEPHGTQGST 217
 Qy 166 PACWCGDPSIDWLLAERAGNAVLSVNPILLIFVEGVQSYN-----GDSY--W 212
 Db 218 RAKWDGTTDKDNFKHFAETASKKILAINPNWLVFVGVETLTPKGVPTWSTGLTDYCTW 277
 Qy 213 WGNLQAGQYPPVVLNV-PNRLVYSADHYATSVGPQTFW-SDTPFPNNMGINWKNWGYL 270
 Db 278 WGNLGRGVDRDHPIDLAGHQDQLVYSPHDYGLVFDKWFQKDFDKASLTADVWGNMPLFI 337
 Qy 271 FNQNIAPVWLGEGFTTL-QSTTDQW---LKTLYQLRPTAQYGADSFQWTFWSPDSDG 326
 Db 338 HDEDIAPLLIGEWGRLGQDPQDKWMAALRDLVAERR-----LSQTFWVLNPNMSG 388
 Qy 327 DTGGILKDDWQTVDTVK 343
 Db 389 DTGGLLDDWKTWDEVK 405
 RESULT 7
 GUNB_CLOTM STANDARD; PRT; 563 AA.
 ID GUNB_CLOTM
 AC P04956;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.3333 Seconds
(without alignments)
2236.934 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987

Sequence: 1 AGGGYWHTSREILDANNVP.....VDTKDYLAIPKSIIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1978	99.5	562	1 GUN1_ACICE	P54583 acidothermu
2	1052	52.9	397	1 GUN1_FABPO	P23548 paenibacill
3	800.5	40.3	484	1 GUN4_XANCP	P19487 xanthomonas
4	824.5	31.4	1039	1 GUNB_CALSA	P10474 c endogluc
5	600.5	30.2	566	1 GUNB_CLOTM	Q05352 clostridium
6	549.5	27.7	747	1 GUNB_CELFI	P50400 cellulomona
7	522	26.3	563	1 GUNB_CLOTM	P04956 clostridium
8	291.5	14.7	748	1 GUNB_PSEFL	P27033 pseudomonas
9	187	9.4	424	1 GUN1_RALSO	P58599 ralstonia s
10	175	8.8	426	1 GUN2_FALSO	P17974 ralstonia s
11	161.5	8.1	814	1 GUNB_CLOTM	P10477 clostridium
12	147.5	7.4	406	1 GUN1_RUMAL	P16216 ruminococcu
13	140.5	7.1	584	1 GUNB_CLOCE	P25472 clostridium
14	137	6.9	900	1 GUNH_CLOTM	P16218 clostridium
15	136.5	6.9	388	1 GUN3_HUMIN	Q12624 humicola in
16	136	6.8	409	1 GUNB_RUMAL	P23661 ruminococcu
17	132	6.6	438	1 EXG_CANAL	P29717 candida alb
18	128	6.4	448	1 GUN_CLOSA	P15704 clostridium
19	126	6.3	825	1 GUN3_BACSA	P19570 bacillus sp
20	122.5	6.2	517	1 GUNA_CLOLO	P54937 clostridium
21	120.5	6.1	658	1 GUN3_FIBSU	P14250 fibrobacter
22	120	6.0	336	1 GUNA_RUMFL	P16169 ruminococcu
23	120	6.0	440	1 GUNB_CLOCL	P28621 clostridium
24	119	6.0	445	1 SPRI_YEAST	P32603 saccharomyc
25	117.5	5.9	505	1 GUNV_ERWCA	Q47096 erwinia car
26	117	5.9	501	1 YBQ6_YEAST	P38081 saccharomyc
27	115.5	5.8	504	1 GUNW_ERWCA	Q59385 erwinia car
28	113.5	5.7	418	1 GUNW_ERWCA	P07982 trichoderma
29	113	5.7	444	1 GUNN_ERWCA	Q59394 erwinia car
30	112.5	5.7	429	1 GUNA_BUTFI	P22541 butyrivibri
31	112.5	5.7	532	1 HEXA_DICDI	P13723 dictyosteli
32	112	5.6	764	1 YIS7_YEAST	P40566 saccharomyc
33	111	5.6	562	1 EXG2_YEAST	P52911 saccharomyc

34	110	5.5	515	1 GUNB_CLOCL	P28623 clostridium
35	109.5	5.5	475	1 GUNA_CLOCE	P17901 clostridium
36	108.5	5.5	459	1 GUNA_STRLI	P27035 streptomyce
37	108.5	5.5	466	1 GUN5_THRFU	Q01786 thermomonos
38	107.5	5.4	800	1 GUN_BACSI	P06564 bacillus sp
39	107	5.4	312	1 GUN4_RUMAL	Q07940 ruminococcu
40	106.5	5.4	499	1 GUN2_BACSU	P10475 bacillus su
41	105.5	5.3	522	1 MAL2_DROME	P07190 drosophila
42	105.5	5.3	547	1 GUN1_BUTFI	P20847 butyrivibri
43	104.5	5.3	364	1 GUNA_RUMAL	P23660 ruminococcu
44	104.5	5.3	499	1 GUN3_BACSU	P23549 bacillus su
45	104	5.2	1024	1 SYIP_STAAU	P41368 staphylococ

ALIGNMENTS

RESULT 1

ID	GUN1_ACICE	STANDARD;	PRT;	562 AA.
AC	P54583;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)			
DE	(Cellulase E1) (Endocellulase E1).			
OS	Acidothermus cellulositicus.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Frankineae; Acidothermaceae; Acidothermus.			
OX	NCBI_TaxID=28049;			
RN	[1]_TaxID=28049;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 43068 / 11B;			
RA	Layman R.A., Himmel M.E., Thomas S.R.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.			
RX	MEDLINE=96346058; PubMed=8718854;			
RA	Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplus P.A.;			
RT	"Crystal structure of thermostable family 5 endocellulase E1 from			
RT	Acidothermus cellulositicus in complex with cellotetraose.";			
RL	Biochemistry 35:10648-10660(1996).			
CC	FUNCTION: Has a very high specific activity on			
CC	carboxymethylcellulose.			
CC	CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.			
CC	CELLULAR: Thermotable enzyme with an optimal temperature of			
CC	81 degrees Celsius.			
CC	FUNCTION: Belongs to cellulase family A (family 5 of glycosyl			
CC	hydrolases).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U33212; AAA75477.1; -			
DR	PDB; 1BCE; 14-OCT-96.			
DR	PDB; 1C0D; 23-JUL-99.			
DR	InterPro; IPR001919; Bac celose-bind.			
DR	InterPro; IPR008965; Cellul bind.			
DR	InterPro; IPR001547; Glyco_hydro_5.			
DR	Pfam; PF00553; CBM 2; 1.			
DR	Pfam; PF00150; cellulase; 1.			
DR	SMART; SM00637; CBD; II; 1.			
DR	PROSITE; PS00859; GLYCOSYL_HYDROL_F5; 1.			
DR	Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.			
FT	SIGNAL 1 41			
FT	CHAIN 42 562			
FT	DOMAIN 42 400			
FT	CATALYTIC.			

This Page Blank (uspto)

QY	68	DIILKPGTSPNSINFIYQ-----NNODLQG-----LTSLOVMDKIVAYAGQIGLRI	111
Db	66	DLMTNETLANNVTVRSQFSQSLGNDIDIVGQTNPNPSIIDLPLIEAYKTVVTTILGNNDMV	125
QY	112	ILDRH--RPD--CSGQ--SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD	165
Db	126	ILDNHLTKPGWCCANDGNGFFGQDFDPTVWAAUKMAATFNGVSNVVGMSLRNELRG	185
QY	166	PACWGCDDPSIDWRLAARAGNAVLSVNPILLIFVEGVQSYNGDSYWGNGLQAGQYPV	225
Db	186	PK-----QVNDWFKYMQOGAEAVHSANKVILSLG--SFDAD-----LSFVSRPV	232
QY	226	VLVNPNRLVYSAHYATYSGPQWFSPTPPNMPG--IWNK---NWGYLFNQNIAPVWL	281
Db	233	KLSFTGKLVFELHWYSPSDG--NSWAANN--PNDICGVLNRI GNGGYLLNQGF--PLFLS	288
QY	282	EFGTTLQ--STTDQTLKTLVQYLRPTAQYGADSFQWTFW-----SNPDSGDTG-----G	330
Db	289	EFIDERGVTNDRYFGCL-----TGWAAENDVDWSLWALTCSYYLRQKVGMEYYG	342
QY	331	ILKDDWQTV	339
Db	343	VLDSDWISV	351

Search completed: August 2, 2004, 16:37:37
Job time : 13.3333 secs

QY 330 -GILKDDWQTV 339
Db 365 YGMLDANWHNV 375

RESULT 15
T51476
hypothetical protein K3M16_70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51476
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T51476
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-526 <SAT>
A:Cross-references: EMBL:AL391150
A:Experimental source: cultivar Columbia; BAC clone K3M16
C:Genetics:
A:Map position: 5
A:Introns: 135/3; 218/2; 327/2; 393/1
A:Note: K3M16_70

Query Match 11.1%; Score 221; DB 2; Length 526;
Best Local Similarity 23.5%; Pred. No. 1.6e-09;
Matches 87; Conservative 60; Mismatches 160; Indels 64; Gaps 14;

QY	8	TSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLPY--	65
Db	30	TKSRWIVNNKGRVYKACANW--PSHLKPVVAEGLSSQPMDSISKIKDMGFNCVRLTWPL	88
QY	66	-----SDDILKEPTMPNSINFQMNQDLGLTS-----LQVMDKIVAYAGQIGLRI	111
Db	89	ELMINDTLAFNVTVKQSFERYGLDHELGQYTHNPYIWTPLINVOAVVYSLGRHDVMV	148
QY	112	ILDRHRP-----DCSGQ--SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD	165
Db	149	ILDNHLTKPGWCCANDGNGFFGQDFDPTVWAAUKMAATFNGVSNVVGMSLRNELR	207
QY	166	PACWGCDDPSIDWRLAARAGNAVLSVNPILLIFVEGVQSYNGDSYWGNGLQAGQYPV	225
Db	208	-----GYNHTSKDYKYMKGAEAVHTSNPNVLISGLN-----FDADLSFLKDRPV	255
QY	226	VLVNPNRLVYSAHYATYSGPQWFSPTPPNMPG--IWNK---NWGYLFNQNIAPVWL	283
Db	256	NLSFKKLIVLELHWYSFTDGTGQWKNHNVDFCSQMFSEKERRTCGGFVLDQGF--PLFLSEF	314
QY	284	GT-----TLQSTTDQTLKTLVQYLRPTAQYGAD--SFQWTFWS-----WNPDSGDTG-----	329
Db	315	GTQQRGGDLGN-----RYMNCMLAAEAEKDLDAWVAWTVGYFYREGKRGVVEA	364

[illegible]

```

Db      :          :          :          :          :          :          :          :          :          :
404    FIKLADQNDIQITIDIHSCSNYGVWRAGELDARPPVDANRVGYDTFREBYSCSATNPS   463

Qy      :          :          :          :          :          :          :          :          :          :
134    EAT-----WISLOALA--QRKGNPTVVGGEDLNHPHDPAACGCGDSPSDWLRAA   182

Db      :          :          :          :          :          :          :          :          :          :
464    SVTRIHYDQKWLANREJLAGLSAKLYSGNLIGDIVFNPYD-YTWA-----EWKGWV   516

Qy      :          :          :          :          :          :          :          :          :          :
183    BRAGNAVLSNPNLLIFVEGVQSYGDSY-----WGNGLOGAQCY   223

Db      :          :          :          :          :          :          :          :          :          :
517    BEAYCAINEVPNNMLIIVEGI-SANANTODGTPD*SVPPHGSGTDLPNPGWENIYEAGAN   575

Qy      :          :          :          :          :          :          :          :          :          :
224    PWLINVP-NRLVYSAHDYATSVPQTWFSDPT-----FPENNMPG   261

Db      :          :          :          :          :          :          :          :          :          :
576    PP--NI PKDRLLFSPTHYGSPVFQQPDPQAQTECAGLEGDEAAQAACRIVINPTVLEQ   633

Qy      :          :          :          :          :          :          :          :          :          :
262    IWNKNWGLFNQNIAPVWLGEFG-----TTLOSTDOTWLTVOYL   303

Db      :          :          :          :          :          :          :          :          :          :
634    GWESHGYLRELGYG-ILIGEFGNMDWPGAKSQADRNAWSHITNTVDQQOQQAASYP   692

Qy      :          :          :          :          :          :          :          :          :          :
304    RPTAQYGADSFQMTFWSWNPDSGDTGGILKDWMQTVDTVKDGYLAPIKSIPDPV   358

Db      :          :          :          :          :          :          :          :          :          :
693    X---PKGINA---CYWSMNPESADTWG---W-----YLFPE-----WDPV   722

```

RESULT 14

E86265

protein F3F19.15 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86265

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.B.; Creasy, T.H.; E.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 815-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Author: Schwarz, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E86265

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-522 <STO>

A:Cross-references: GB:AEC005172; NID:G4850396; PID:RAD31066.1; GSFDB:GNC00141

C:Genetics:

A:Gene: F3F19.15

A:Map position: 1

```
Query Match      13.0%; Score 258.5; DB 2; Length 522;
Best Local Similarity 25.2%; Pred.No. 1.ee-12;
Matches 93; Conservative 62; Mismatches 153; Indels 61; Gaps 18
```

```
QY      8 TSGREILDANNPVRIAGINWGFETCYVYHGLWRDYSRLDDQIKSLGYTIRLPSTD 67
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       7 TSSRWIVDENLGRVKLVCAWN--PSHLQFVAEGLSKQFDVAKKIIVEMGNCVRITWPL 65
```

Db 343 YAQWFEKFTMQTLDDYVDYTWAYTHDQGIAPILIGWGHGMDGKQKQWTLRLDYI 402
 Qy 304 RPTAQYADSFQWTFWNSPDGDTGGILKDDQWTVTKDGYLAP 349
 Db 403 -----VONRIHHTFWCINENSGDTGGLGNDWSTWDEAKYALKP 442

RESULT 10
 B47093
 Cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi
 N;Alternate names: endo-1,4-beta-glucanase
 C;Species: Cellulomonas fimi
 C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
 R;Accession: B47093
 R;Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
 J. Bacteriol. 175, 1910-1918, 1993
 A;Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),
 A;Reference number: A47093; MUID:93209933; PMID:8458833
 A;Accession: B47093
 A;Status: preliminary
 A;Molecule type: DNA; protein
 A;Residues: 1-747 <MEI>
 A;Experimental source: ATCC 484
 A;Note: sequence inconsistent with the nucleotide translation
 A;Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIPI:128122)
 C;Function:
 A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A;Pathway: cellulose degradation
 C;Superfamily: bacterial cellulose-binding domain homology
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation
 F;644-745/Domain: bacterial cellulose-binding domain homology <BCB>
 F;645-744/Disulfide bonds: #status predicted

Query Match 27.7%; Score 549.5; DB 2; Length 747;
 Best Local Similarity 35.0%; Pred. No. 5.3e-35;
 Matches 132; Conservative 58; Mismatches 142; Indels 45; Gaps 15;
 Qy 1 AGGYSW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSLMDQIKSLGYN 59
 Db 40 ATGDDWLHVEGNTIVDTGKAILSGVNVWFGFNASERVFHGLWSGNITQTQMAQRGIN 99
 Qy 60 TIRLPYSDDIL---KPGTM--PNSINFYQNNQDLQGLTSIQVMDKIVAYAGQIGLRIILD 114
 Db 100 VVRYPVSTQLLEKWKAGTFLKPN-VNTY-ANPELEGKNSIQIFEYWLTLCKQYGIKVELD 157
 Qy 115 RH--RPPCSGQ-SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPH-----D 165
 Db 158 VHSADNSGHVYNNMWKGDITTEDVYEGWEWAATRWKDDDTIVGADIKNEPHGTQGSTE 217
 Qy 166 PACGCGDPSIDWLAERAGNAVLSVNPILLIFVEGVQSYN-----GDSY--W 212
 Db 218 RAKWDGTTDKNFHFAETASKILAINPNWLVEGEVIEYPRGVPWTSGTLDYGIW 277
 Qy 213 WGNLQAGQYPPVLNV-PNRLVYSAHDYATSVGQPTWP-SDTPFNPNMGINWKNWGYL 270
 Db 278 WGNLGRVDRHPIDLGAHQDLQVLYSPHDYGLVFDQKWFQKDFDKASLTADVWGNPNLFI 337
 Qy 271 FNQNIAPVWLGEGFTTI-QSTDTQW---LKTAVOYLRLPTAQYCADSFQWTFWNSPDG 326
 Db 338 HDEIAPLLIGEWGRGLQDPQDKQKWAALRDILVAER-----LSQTFWVLNPNNG 388
 Qy 327 DTGGLKDDQWTVTK 343
 Db 389 DTGGLLDDWKTDVVK 405

RESULT 11
 CZCLBM
 Cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum
 N;Alternate names: endo-1,4-beta-glucanase B precursor
 C;Species: Clostridium thermocellum
 C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 18-Jun-1999

C;Accession: A23512
 R;Grepinet, O.; Beguin, P.
 Nucleic Acids Res. 14, 1791-1799, 1986
 A;Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endoglucan
 A;Reference number: A23512; MUID:86148508; PMID:3453102
 A;Accession: A23512
 A;Molecule type: DNA
 A;Residues: 1-563 <GRE>
 A;Cross-references: GB:X03592; NID:g40668; PIDN:CAA27266.1; PID:g40669
 C;Comment: This secretory enzyme is part of a highly active and thermostable cellulase co
 C;Genetics:
 A;Gens: celB
 C;Function:
 A;Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans
 A;Pathway: cellulose degradation
 C;Superfamily: cellulase B; Clostridium cellulase repeat homology
 C;Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de
 F;1-25/Domain: signal sequence #status predicted <SIG>
 F;26-563/Product: cellulase B #status predicted <MAT>
 F;502-525/Domain: Clostridium cellulase repeat homology <CCR1>
 F;534-557/Domain: Clostridium cellulase repeat homology <CCR2>
 Query Match 26.3%; Score 522; DB 1; Length 563;
 Best Local Similarity 30.7%; Pred. No. 5.1e-33;
 Matches 129; Conservative 58; Mismatches 159; Indels 74; Gaps 13;
 Qy 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSM 50
 Db 27 AEGSYADLAEPDDDDWLHVEGNTIVDKYGNKVIWGANNWFGNCRERMLDSDYHSDIADI 86
 Qy 51 DOIKSLGYNITRLPYSDIL---KPGTMPSINFYQNNQDLQGLTSIQVMDKIVAYAGQI 107
 Db 87 ELVADKGINVVMPIATDLYAWSGIYPPSTDISYNNPALAGLSVLEFNFMLENPKRV 146
 Qy 108 GLRIILDRHDPDCSQS---ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPH 164
 Db 147 GKVLVDHSPETDQGHNYPLWYNTTITEIFKAWVWVAERYKNDTTIIGFDLKNPEH 206
 Qy 165 -----DPACGCGDPSIDWLAERAGNAVLSVNPILLIFVEGVQSY----- 206
 Db 207 TWGTMTKIAQSAIWDDSNHNNKRVAAETALALEVHPNVLIFVEGVEMYPKDGWDD 266
 Qy 207 -----NGDSY--WGNLQAGQYPPVL-NVPNRLVYSAHDYATSVGQPTWSPDP 253
 Db 267 EFTDTPWNTGNDYGNWGNLGRGVDPINLGYSQLVSPHDYGPVYVEQDWFKGD 326
 Qy 254 TFPNN-----MPGINWKNWGYLFNQNIAPVWLGEGFTTIQSTDTQWTLKTLVOYLRLP 305
 Db 327 FTTANDEQAKILYEQCRDNRNWAYIMEEGISPLLLGEWGMTEG--GHPLDLNLKYLRC 384
 Qy 306 TAYQAGDS---FQWTFWNSPDGDTGGILKDD-----WQTVDTVKDGYLAPI 350
 Db 385 MEDFILENKYKLHHTFWCINIDSADTGLFTRDEGTPPGGRDLKWN--DNKYDNYLYPV 442
 RESULT 12
 B82761
 Extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a5c
 C;Species: Xylella fastidiosa
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C;Accession: B82761
 R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; MUID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: B82761
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-565 <SIM>
 A;Cross-references: GB:AE003920; GB:AE003949; NID:g9105697; PIDN:AAF83620.1; GSPDB:GN0012
 A;Experimental source: strain 9a5c
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
Query Match 31.4%; Score 624.5; DB 2; Length 915;
Best Local Similarity 35.5%; Pred. No. 9.6e-41;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSRLMDQIKSLGYNTIRLP 64
DB 505 WLYVSGNKIVDXDGRFVWLTGINWFGYNTGTNVFDGWSNCLKDTLAEIANRGEFLLRVP 564

QY 65 YSDIL---KPGTMEN-SINFYQMNDLQGLTSQVMQKIVAYAGQIGLRIILDRH--RP 118
DB 565 ISAEILLNWSQGIYKPKPNINY-VNPELEGRKNSLEVFIWQTCKEVGKIMLDIHSIKT 623

QY 119 DCSGO-SALWYTSVSEATWISDLOALAQRYKGNPTVWGFDELHNEPHDP-----ACWG 170
DB 624 DAMGHIYPVWYDEKFTPEDFYKACEWITNRYKXNDTIIAFDLKNEPHGKFWODTTFAKWD 683

QY 171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWVGNNL 217
DB 684 NSTDINNWKYAAETCAKRIILINPNLLIVIEGIEAYPKDDVTWTSKSSSDYSTWVGNNL 743

QY 218 QGAGQYPPVL-NVPNRLVYSAHDYATSGPOTWFSDPFPNN--MPGIWNKNGYLFNQ 274
DB 744 RGVKRPINLKGKYNKYVSPHDYGPVYQQPWYF-PGFTKESLLQDCWRPNWAYIMEEN 802

QY 275 IAPVWLGEFGTTLQSTTDQTLWKLTVQYLRPTAQYGADSFQWTFWSNPNDSGDTGILKD 334
DB 803 IAPLLIGWGGHLDGADNEKWKYLRDYII-----ENHIHTFCFNANSODTGLGVY 856

QY 335 DWQTVTVKDGYLAP 349
DB 857 DFTTWDEKYSFLKP 871

RESULT 8
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C:Species: Caldocellum saccharolyticum
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: S02711
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chanley, L.W.; Bergquist, P.L.
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo-1,4-beta-glucanase
A:Reference number: S02711; MUID:89098398; PMID:2789517
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1-1039 <SAU>
A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C:Genetics:
A:Gene: celB
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>
F:29-1039/Product: cellulase #status predicted <NAT>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
Query Match 31.4%; Score 624.5; DB 2; Length 1039;
Best Local Similarity 35.5%; Pred. No. 1.1e-40;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSRLMDQIKSLGYNTIRLP 64
DB 629 WLYVSGNKIVDXDGRFVWLTGINWFGYNTGTNVFDGWSNCLKDTLAEIANRGEFLLRVP 688

QY 65 YSDIL---KPGTMEN-SINFYQMNDLQGLTSQVMQKIVAYAGQIGLRIILDRH--RP 118
DB 689 ISAEILLNWSQGIYKPKPNINY-VNPELEGRKNSLEVFIWQTCKEVGKIMLDIHSIKT 747

QY 119 DCSGO-SALWYTSVSEATWISDLOALAQRYKGNPTVWGFDELHNEPHDP-----ACWG 170
DB 748 DAMGHIYPVWYDEKFTPEDFYKACEWITNRYKXNDTIIAFDLKNEPHGKFWODTTFAKWD 807

QY 171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWVGNNL 217
DB 808 NSTDINNWKYAAETCAKRIILINPNLLIVIEGIEAYPKDDVTWTSKSSSDYSTWVGNNL 867

QY 218 QGAGQYPPVL-NVPNRLVYSAHDYATSGPOTWFSDPFPNN--MPGIWNKNGYLFNQ 274
DB 868 RGVKRPINLKGKYNKYVSPHDYGPVYQQPWYF-PGFTKESLLQDCWRPNWAYIMEEN 926

QY 275 IAPVWLGEFGTTLQSTTDQTLWKLTVQYLRPTAQYGADSFQWTFWSNPNDSGDTGILKD 334
DB 927 IAPLLIGWGGHLDGADNEKWKYLRDYII-----ENHIHTFCFNANSODTGLGVY 980

QY 335 DWQTVTVKDGYLAP 349
DB 981 DFTTWDEKYSFLKP 995

RESULT 9
A40589
cellulase (EC 3.2.1.4) - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium thermocellum
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jun-2000
C:Accession: A40589; S31381
R:Lemaire, M.; Beguin, P.
J: Bacteriol. 175, 3353-3360, 1993
A:Title: Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri;
A:Reference number: A40589; MUID:93273701; PMID:8501039
A:Accession: A40589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <LEM>
A:Cross-references: GB:X69390; NID:g40677; PIDN:CAA49187.1; PID:g40678
C:Genetics:
A:Gene: celG
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A:Pathway: cellulose degradation
C:Superfamily: cellulase B; Clostridium cellulase repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:503-526/Domain: Clostridium cellulase repeat homology <CCR>
F:536-559/Domain: Clostridium cellulase repeat homology <CCR>
Query Match 30.2%; Score 600.5; DB 2; Length 566;
Best Local Similarity 34.2%; Pred. No. 3.8e-39;
Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSRLMDQIKSLGYNTIRLP 64
DB 44 WLCCKGNKIYDMYGNVWLTGANWFGFNCSECFHGAW-YDVKTILTSIADRGINLRIP 102

QY 65 YSDIL---KPG-----TMPNSINFYQMNODLQ-----LTSLOWVMDKIVAYAGOI 107
DB 103 ISTELLYSWMIGKPNVSVTASNPPYHVNVDFYDPETDDVKNSEWIFDIIMGCKEL 162

QY 108 GLRIILDRHPPDC--SGQS-ALWY---TSS---VSEATWISDLOALAQRYKGNPTVWGF 158
DB 163 GIKVMIDIHSPDANNSGHVYELWYKETSCTGVVTTKMWIDTLVWLADKYKNDTIIAFD 222

QY 159 LHNEPHDP-----ACWCGGSPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSY- 206
DB 223 LKNEPHGKRGYTAEVFKLLAKWNSDNNKNTAETCAKAILLEVAPKVLIVIEGVQYP 282

QY 207 -----NGD-----SYWVGNNLQAGQYPPVLLNPN--RLVYSADHYATSV 244
DB 283 KTEKGYTYDTPDIWGTGADSPWYSAWVGNNLKGVDYPIDLGLPNSQIVYSPHDYGPSV 342

QY 245 GPQTWP-SDTTFPNNMPGIWNKNGYLFNQIAPVWLGEFGTTLQSTTDQTLWKLTVQYL 303

Db 321 ALVDYL---ISIGVTDTF---YWSINANSSDGTGGLFRDDMNHV 357

RESULT 6
E75142
endoglucanase PA80632 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: E75142
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A;Reference number: A75001
A;Accession: E75142
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-514 <KAW>
A;Cross-references: GB:AU248285; GB:AL096836; NID:G5458067; PIDN:CAB49854.1; PID:G5458366
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: celB-like; PAB0632
C;Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 34.3%; Score 682.5; DB 2; Length 514;
Best Local Similarity 41.6%; Pred. NO. 1.3e-45;
Matches 148; Conservative 57; Mismatches 122; Indels 29; Gaps 10;

QY 7 HTSGREIIDANNV-----PVRIAGINWFGTETCNVVHGLWSRDYRSMLDQIKSLGYNT 60
DB 30 YTAENGIIIFVNVTGTGEKKPLVHGVSWFGLXDHVVYGLDKENWKDLKDXVLGFNA 89
QY 61 IRLPYSDIILPQTNP--SINFQMNOODLGTLSQLMKIVAYACIGLRILIILDRHP 118
DB 90 IRLPCSEIRIPDRPSPERIN-YELNDPLKNLTSLTEIMEXIIIEANSIGLYILLDYHI 148
QY 119 DCSGOSALWTSSVSEAATWISDLOALAQRYKGNPTVVGFDLHNPHPPAPCWGCCDPDSIW 178
DB 149 GCETIEPLWYTENTYSEEQYIKDWIFLAKRCKYPNVIGADIKNEPHGEAGWGTDGER-D 207
QY 179 RLAAERAGNAVLSNPNNLLIIVEGVQ-----SYNGDSYW--WGNLOGACQYPVVL 227
DB 208 RLFAEKVGREILKVAPHWLIFVEGTQYTHVPNIDEILEKKGWTFWGENLMGVKDYVR 267
QY 228 NVPNRLVSAHDYATSUGPQWFSDPPTPNMPGWNKMGYLFNQNIAPVWLGEFGTTL 287
DB 268 -PRGVVYSPIHVGSVVMMDYFKSPDFNNPMIIEWTHFGYLDLNVTLV-IQEWGNT 325

QY 288 OSTTDQTWLKTLVQLRPTAOGADSFQWTFWSNPNDSGGDTGGILKDWCQTVDRVK 343
 : | | | | | | | |
DB 326 EG-LBKWQDAFVKWLKKIY-----NFFWVCINPESGDTGGIFLFDWKVTANWEK 375

RESULT 7

A43802 cellulase [EC 3.2.1.4] / cellulose 1,4-beta-cellobiosidase [EC 3.2.1.91] - Caldocellum se
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Caldocellum saccharolyticum
C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C:Accession: A43802
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "C"
A:Reference number: A43802; MUID:91136262; PMID:2126700
A:Accession: A43802
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual transla
A:Molecule type: DNA
A:Residues: 1-915 <SAU>
A:Cross-references: EMBL:X13602
C:function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

JH0158
cellulase (EC 3.2.1.4) precursor - Xanthomonas campestris pv. campestris
N:Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase
C:Species: Xanthomonas campestris pv. campestris
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: JH0158
R:Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.
Gene 89, 53-59, 1990
A:Title: Nucleotide sequence of the engXCA gene encoding the major endoglucanase of Xanthomonas campestris pv. campestris
A:Reference number: JH0158; MUID:90323605; PMID:2373365
A:Accession: JH0158
A:Molecule type: DNA
A:Residues: 1-493 <GOU>
A:Cross-references: GB:M32700; NID:gl55397; PIDN:AAA27612.1; PID:gl55398
C:Comment: Endoglucanase plays a minor role in the early stages of pathogenicity of Xanthomonas campestris pv. campestris
C:Genetics:
A:Gene: engXCA
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A:Pathway: cellulose degradation
C:Superfamily: Xanthomonas campestris cellulase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-474/Product: cellulase #status predicted <MAT>
F:375-399/Region: proline/threonine-rich

Query Match 40.3%; Score 800.5; DB 1; Length 493;
Best Local Similarity 47.2%; Pred. No. 7.2e-55;
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSLMDQIKSLGYNTIRLP 64
Db YSINNSRQIVDSDGKVVQLKGVNFGFETGNHVMHGLWARKMDIMVQMGFGFNAVLRLP 85
QY 65 YSDILKPTGMPNSINFYQMNQDLQGLTSLOVMDKIVAVAGQIGLRIILDRHPCSGQS 124
Db FCPATLRSDTMFASID-YSRNADLQGLTSLOVMDKIVAVAGQIGLRIILDRHPCAGIS 144
QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPAWCGDPSIDWFLAAR 184
Db ELWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPAWCGDPSIDWFLAAR 204
QY 185 AGNAVLNPNLLIFVEGVQ-----SYNGDSYVWGNLQAGQYVPLNVP-NRLVYSAH 238
Db GSAAVLAVAPKWLIAVEGTDNPNVCTNG-GIFWGNLQPLACTP--LNIPANRLLALPH 261
QY 239 DYATSVGPOTWSDPTFPNNMFGIWNKMGYLFNQNIAPVWLGEGTTL--QSTTDQWL 296
Db VYGPVDFVQSYFNDNFPNNMFAIWEHFGQAGTH--ALLGEGGYGEGDARDKTWQ 319
QY 297 KTLVQLRRTAQYGADSFQWTFWNPDSGDTGGILKDDQTV 339
Db DALVKYLR---SKGIN--QGFYWSWNPNSGDTGGILRDDWTSV 357

RESULT 3
E71059
probable endo-1,4-beta-glucanase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: E71059
R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71059
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-458 <KAW>
A:Cross-references: GB:AP000005; NID:G3236132; PIDN:BAA30271.1; PID:G3257598
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank

JH0158
cellulase (EC 3.2.1.4) precursor - Xanthomonas campestris pv. campestris
N:Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase
C:Species: Xanthomonas campestris pv. campestris
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: JH0158
R:Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.
Gene 89, 53-59, 1990
A:Title: Nucleotide sequence of the engXCA gene encoding the major endoglucanase of Xanthomonas campestris pv. campestris
A:Reference number: JH0158; MUID:90323605; PMID:2373365
A:Accession: JH0158
A:Molecule type: DNA
A:Residues: 1-493 <GOU>
A:Cross-references: GB:M32700; NID:gl55397; PIDN:AAA27612.1; PID:gl55398
C:Comment: Endoglucanase plays a minor role in the early stages of pathogenicity of Xanthomonas campestris pv. campestris
C:Genetics:
A:Gene: engXCA
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A:Pathway: cellulose degradation
C:Superfamily: Xanthomonas campestris cellulase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-474/Product: cellulase #status predicted <MAT>
F:375-399/Region: proline/threonine-rich

Query Match 40.3%; Score 800.5; DB 1; Length 493;
Best Local Similarity 47.2%; Pred. No. 7.2e-55;
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSLMDQIKSLGYNTIRLP 64
Db YSINNSRQIVDSDGKVVQLKGVNFGFETGNHVMHGLWARKMDIMVQMGFGFNAVLRLP 85
QY 65 YSDILKPTGMPNSINFYQMNQDLQGLTSLOVMDKIVAVAGQIGLRIILDRHPCSGQS 124
Db FCPATLRSDTMFASID-YSRNADLQGLTSLOVMDKIVAVAGQIGLRIILDRHPCAGIS 144
QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPAWCGDPSIDWFLAAR 184
Db ELWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPAWCGDPSIDWFLAAR 204
QY 185 AGNAVLNPNLLIFVEGVQ-----SYNGDSYVWGNLQAGQYVPLNVP-NRLVYSAH 238
Db GSAAVLAVAPKWLIAVEGTDNPNVCTNG-GIFWGNLQPLACTP--LNIPANRLLALPH 261
QY 239 DYATSVGPOTWSDPTFPNNMFGIWNKMGYLFNQNIAPVWLGEGTTL--QSTTDQWL 296
Db VYGPVDFVQSYFNDNFPNNMFAIWEHFGQAGTH--ALLGEGGYGEGDARDKTWQ 319
QY 297 KTLVQLRRTAQYGADSFQWTFWNPDSGDTGGILKDDQTV 339
Db DALVKYLR---SKGIN--QGFYWSWNPNSGDTGGILRDDWTSV 357

RESULT 3
E71059
probable endo-1,4-beta-glucanase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: E71059
R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71059
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-458 <KAW>
A:Cross-references: GB:AP000005; NID:G3236132; PIDN:BAA30271.1; PID:G3257598
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:
A:Gene: PH171
C:Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 38.0%; Score 755; DB 2; Length 458;
Best Local Similarity 45.0%; Pred. No. 2.4e-51;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSLMDQIKSLGYNTIRLPYS 67
Db TSCEE-----TPIHLFGVNVWFGFETPNVHVLGKRWKRMEDMLLQIKSLGFNAIRLPFC 107
QY 68 DILKPTGMPNSINFYQMNQDLQGLTSLOVMDKIVAVAGQIGLRIILDRHPCSGQSALW 127
Db ESKAPGQPTGID-YKNPDLRLDLSQIMEKLIKAGDLGIFVLLDYHRIAGTHIEPLW 166
QY 128 YTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db YTEDFSSEDFINTWIEVAKFGYVNVIGADLKNEPHTSVTPPAAYTDGTGATWGMGNA 226
QY 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYVWGNLQAGQYPVV 226
Db TDWNLAAERIGKAILKVAPHWLIIFVGTQPTNPKTSSYKMGYNWAGGMLMAVKDYPV- 285
QY 227 LNVP-NRLVYSAHDYATSVGPOTWFSIPT-PPNNMFGIWNKMGYLFNQNIAPVWLGEEFG 284
Db NLPRNKLIVSPHYGPDVYNQYFGPAKGFDPNLDPDIWYHFGYVVKLELGYSVVIGFEG 344
QY 285 TTL---QSTTDQWLKTLVQLRRTAQYGADSFQWTFWNPDSGDTGGILKDDQTV 339
Db GYGHGGDPRDVIWONKLVDMW--IENKFCDF--YWSWNPDSGDTGGILODDWTI 397

RESULT 4
E97012
probable non-processive endoglucanase family 5, secreted, Cela homolog secreted, dockerir
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: E97012
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78888.1; PID:gl5023812; GSPDB:GNO0168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0912

Query Match 34.7%; Score 690; DB 2; Length 482;
Best Local Similarity 37.9%; Pred. No. 3e-46;
Matches 141; Conservative 71; Mismatches 126; Indels 34; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSLMDQIKSLGYNTIRLP 64
Db YLSHSDGSKLLDDYGNQVRMTGIAWFGLETENYCFHGLWANRLNINLVADNGFNTLRVP 83
QY 65 YSDILK---KPGT--MPNSINFYQMNQDLQGLTSLOVMDKIVAVAGQIGLRIILDRH 119
Db LSELVNWQWQGVYPTFDSINDY-ISPGLKQNSQLDLDVIAYSKVKVWGLMDHRIE 142
QY 120 CSQOSALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPAC-----WGCG 172
Db SGGQTATWYTSKYTTDDYEKCWQVLADRYKNDTDTIAADTFNEPHGKAVRAETSAKWN 202
QY 173 DPSIDWELAAERAGNAVLNPNLLIFVEGVQSY-----NGDSY---WVGNGNLOG 219
Db TDENWRYAEKVKKILIDNPVKLIVVEGVETYPKEGTAAGSTNPDYDYGWVGWGNLRG 262

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds
(without alignments)
2792.154 Million cell updates/sec

Title: US-09-997-504A-10
Perfect score: 1987
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: piri.*
2: piri.*
3: piri.*
4: piri.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1054	53.0	397	2 A35136	cellulase (EC 3.2.1.4)
2	800.5	40.3	493	1 JH0158	cellulase (EC 3.2.1.4)
3	755	38.0	458	2 E71059	probable endo-1,4-probable non-proce
4	690	34.7	482	2 E97012	endo-1,4-beta-gluc
5	687.5	34.6	592	2 E82759	endoglucanase PAB0
6	682.5	34.3	514	2 E75142	cellulase (EC 3.2.1.4)
7	624.5	31.4	915	2 A43802	cellulase (EC 3.2.1.4)
8	624.5	31.4	1039	2 S02711	cellulase (EC 3.2.1.4)
9	600.5	30.2	566	2 A40589	cellulase (EC 3.2.1.4)
10	549.5	27.7	747	2 B47093	cellulase (EC 3.2.1.4)
11	522	26.3	563	1 CZCLBM	extracellular endo
12	328	16.5	565	2 B82761	celloextrinase C
13	291.5	14.7	748	2 S18652	protein F31919.15
14	258.5	13.0	522	2 E86285	hypothetical prote
15	221	11.1	526	2 T51476	cellulase (EC 3.2.1.4)
16	211.5	10.6	488	2 T51502	cellulase (EC 3.2.1.4)
17	175	8.8	426	2 A42649	cellulase (EC 3.2.1.4)
18	161.5	8.1	814	1 CZCLBM	mannan endo-1,4-be
19	160.5	8.1	722	2 H96986	endo-1,4-beta-gluc
20	158.5	8.0	516	2 J80134	endoglucanase fami
21	155.5	7.8	356	2 E93732	cellulase (EC 3.2.1.4)
22	148	7.4	1012	2 B93326	cellulase (EC 3.2.1.4)
23	147.5	7.4	406	2 A43722	cellulase (EC 3.2.1.4)
24	141.5	7.1	329	2 C72216	endoglucanase - Th
25	140.5	7.1	500	2 S22458	cellulase (EC 3.2.1.4)
26	140.5	7.1	584	2 J01229	cellulase (EC 3.2.1.4)
27	137.5	6.9	388	1 JC5461	cellulase (EC 3.2.1.4)
28	137	6.9	370	2 G57001	endoglucanase fami
29	137	6.9	900	2 JH0157	cellulase (EC 3.2.1.4)

cellulase (EC 3.2.1.4)
endoglucanase B -
hypothetical prote
endo-1,4-beta-mann
glucan 1,3-beta-gl
beta-glucanase [im
cellulase (EC 3.2.1.4)
cellulase (EC 3.2.1.4)
endoglycosylcerami
cellulase (EC 3.2.1.4)
endoglucanase 3 (E
cellulase (EC 3.2.1.4)
exo-1,3-beta-gluc
cellulase (EC 3.2.1.4)
glucan 1,3-beta-gl

30 136.5 6.9 388 1 S43920
31 136 6.8 409 2 S12018
32 129 6.5 656 2 AB1843
33 128.5 6.5 669 2 D72278
34 128 6.4 438 2 A47702
35 128 6.4 438 2 T52149
36 128 6.4 448 2 A27631
37 126 6.3 825 2 J50174
38 124.5 6.3 455 2 S16559
39 122.5 6.2 482 2 J73332
40 122.5 6.2 517 2 I40798
41 120.5 6.1 659 2 A33598
42 119 6.0 441 2 A44815
43 119 6.0 445 2 A40639
44 119 6.0 783 2 JC5467
45 117.5 5.9 464 2 T40108

cellulase (EC 3.2.1.4) - Bacillus polymyxa
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus polymyxa
C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 24-Sep-1998
C:Accession: A35136
R:Baird, S.D.; Johnson, D.A.; Seligy, V.L.
J. Bacteriol. 172, 1576-1586, 1990
A:Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase
A:Reference number: A35136; MUID:90170877; PMID:2307659
A:Accession: A35136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <BAI>
A:Cross-references: GB:M33791; GB:M33840
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Superfamily: Xanthomonas campestris cellulase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

ALIGNMENTS

RESULT 1

Query Match	53.0%	Score	1054	DB	2	Length	397
Best Local Similarity	54.5%	Pred. No.	8.1e-75				
Matches	194	Conservative	44	Mismatches	106	Indels	12
Gaps							

QY 4 GYWHTSGREILDANNVPVRIAGINWFGPFCNYYVHGLWSRDYRSMLDQIKSLGNTIRL 63
DB 37 GYHTQGNKIVDESGKEAFAFNGLWFLGLETPTNYTLHGLWSRSMDDMLDQVKEGNLIRL 96
QY 64 PYSDILKPGTWNISNFQVNDLQGLTSLQVMDKIVAVAGQIGRLILDRHPDCSGQ 123
DB 97 PYSNQLFSSSRPDSID-YHKNPDLVGLNPQIWDKLEKAGQGIQIILDRHPGSGGQ 155
QY 124 SALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
DB 156 SELWYTSQYPPESRWISDWKMLADRYKKNPTVIGADLHNEPHGQASWGTGNASTDWRLLAQ 215
QY 184 RAGNAVLSNPNLLTFVEGVQ---SYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSADY 240
DB 216 RAGNAVLSNPNLLTFVEGVQ---SYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSADY 275
QY 241 ATSVGPQWFSDFPTFNMMPGIWNKNGYLFNQIAPVLMGFGF--TTLQSTTDGTWUKT 298
DB 276 GFGVSSQWPNFNDPAFSPNLPALWDQTWGYSKQNIAPVLMGFGF--TTLQSTTDGTWUKT 335
QY 299 LVQYLRPTAQYCADSFQWTFWSNPDSDGTGILKDDQVTDVTKDGYLAPIKSSI 354
DB 336 LVHYI-----CANNLYFTYSLNPNFSGDTGGLLDDTWTNRPQDMLGRMKPV 385

RESULT 2

This Page Blank (uspto)

```

; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9262
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9262

Query Match      33.0%; Score 655.5; DB 15; Length 421;
Best Local Similarity 39.9%; Pred. No. 3.1e-55;
Matches 141; Conservative 60; Mismatches 123; Indels 29; Gaps 12;

Qy 5 YWHTSGREILDANNVPRIAGINWFGTETCNVYVHGLWSRDSMLDQIKSLGYNTIRLP 64
Db 26 YSISHGKVVDDKGN-QIQLRGVNWFGPETGHHVYVGLWARNWKEFITQLQCMGFNAIRLP 84
Qy 65 YSDDILKPGTMPNSINFYQNMNDLQGLTSLQVMDKIYAYAGQIGLRILDRHRPDCSGQS 124
Db 85 FCPANLASNTSPSSID-YSRNPDQLGLSSLLQILDKVVKELSDREMYVLLDHRPDCSAIS 143
Qy 125 ALWYTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
Db 144 ELWHTDSYSEKQWIDDLRFVAHRYANVPGVIGLEVKNEPHGRATWGTGDTDMNTAVEH 203
Qy 185 AGNAVLNVNPLLIFVEGVQSYNGDS-----YMWGGMLOGAGQYPPVVLNVP-NRLVY 235
Db 204 AAAAILEAAPKWVIGIGI-----GENICSTIGHFNGENLEPMDCPT--LKVPADHLLL 257
Qy 236 SAHDYATSVGPQTWFSPTFPNNMGIWNKNWYLFNQNTAPWLGEBFTTL--QSTTDQ 293
Db 258 MPHYVGPDPVYQYFNSPDPFNNMAAIWDKHGFAKAGYA-MAIGFEGKYGEGDPRDI 316
Qy 294 TWLKLTYQLRPTAQYG-ADSFQWTFWSWNPDSGDTGGILKDDWQTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNNDWTTPRDDKVK 363

```

Search completed: August 2, 2004, 16:40:48
Job time : 42.3333 secs

Best Local Similarity 42.6%; Pred. No. 3.7e-58;
Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCTNYVHGLSRDYSRLMDQIKSLGYNTIRLP 64
DB 26 YSISHGKVIDDKGN-QIQLGKGSWFGFETCTNYVHGLSRDYSRLMDQIKSLGYNTIRLP 84
QY 65 YSDILKPGTTPNSINFYQWQDQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 124
DB 85 FCPASLNSNTSPSSID-YNRNPDQLGSLQIMDKVKVRLSDRGIVLMDHHTPDCAALS 143
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPCACGDPSPIDWRLAER 184
DB 144 ELWYTSYSEKQWIDRLFAHRYANVPVGLDKVNEPHGRATWGTGDKTDMNTAVEH 203
QY 185 AGNAVLNVNPNLLIFVEGVQSYNGDS---YWGNGNLQAGQYPPVVLNVP-NRLVYSAHD 239
DB 204 AAAAILAARPKWILGVEGICENPSCSTTGHFWGENLEPMDCTP--LKVPADHLLIHV 261
QY 240 YATSGQFQWFSPTPPNNMGIWNNKQWGLFNQNIAPVWLGEFGTTL--QSTTDQWMLK 297
DB 262 YGPDVYVQYFNSPDPFPNNMAALWDKHFHFAKAGYA-MAIGFEGKYGEGDPRDVAQN 320
QY 298 TLVQYLRPTAQYG-ADSFQWTFWSNPDSDGTGGLKDDWQTV 339
DB 321 ALVDYL---ISIGVTDTF---YASINANSSTGGLFRDDWNVH 357

RESULT 13
US-10-369-493-21618
; Sequence 21618, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21618
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21618

Query Match 34.3%; Score 682.5; DB 15; Length 514;
Best Local Similarity 41.8%; Pred. No. 9.2e-58;
Matches 148; Conservative 57; Mismatches 122; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETCTNYVHGLSRDYSRLMDQIKSLGYNT 60
DB 30 YTAENGIIIFVQVNTTCEKKPLYLHGVSFGEKLDKHHVYVGLDKRNWKDILKDVKRLGNA 89
QY 61 IRLPYSDILKPGTTPNSINFYQWQDQGLTSLQVMDKIVAYAGQIGLRIILDRHRP 118
DB 90 IRLPFCSEIRPDTSPERIN-YELNPDKNLTSLEIMEKIEYANSIGLYLLIDYHRI 148
QY 119 DCSGQALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPCACGDPSPIDW 178
DB 149 GCEIEPLWYNTSEYQIKWIFLAKFQKYPVNVIGADIKNEPHGEAGWTGDER-DF 207
QY 179 RLAAERAGNAVLNVNPNLLIFVEGVQ-----SYNGDSYW--WGNLQAGQYPPVL 227
DB 208 RLFAEKVGREILKVPADHLLIHVFGTGYTHVPNIDEIKKGMWTFWGENLMGVKDPVRL 267
QY 228 NVNRLVYSAHDYATSVGQFQWFSPTPPNNMGIWNNKQWGLFNQNIAPVWLGEFGTTL 287

DB 268 -PRGKVVYSHVYGPSYMMDYFKSPDPNNMPTIETHEGYLTDLNLYTLV-IGEWGNY 325
QY 288 QSTTDDQWMLKPGTTPNSINFYQWQDQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 343
DB 326 EG-LDKYVQDAFVKWLIKXIIY-----NFFYWCINPESGDTGGLFDDWKTVANWEK 375

RESULT 14
US-10-369-493-9473
; Sequence 9473, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9473
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9473

Query Match 33.2%; Score 660.5; DB 15; Length 369;
Best Local Similarity 40.5%; Pred. No. 8.3e-56;
Matches 143; Conservative 56; Mismatches 125; Indels 29; Gaps 12;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCTNYVHGLSRDYSRLMDQIKSLGYNTIRLP 64
DB 26 YSISHGKVIDDKGN-QIQLGKGSWFGFETCTNYVHGLSRDYSRLMDQIKSLGYNTIRLP 84
QY 65 YSDILKPGTTPNSINFYQWQDQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 124
DB 85 FCPANLNSNTSPSSID-YSRNPDQLGSLQILDKVVKELSDRRMYVLLDHRSDCSAIS 143
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPCACGDPSPIDWRLAER 184
DB 144 ELWYTSYSEKQWIDRLFAHRYANVPVGLDKVNEPHGRATWGTGDKTDMNTAVEH 203
QY 185 AGNAVLNVNPNLLIFVEGVQSYNGDS---YWGNGNLQAGQYPPVVLNVP-NRLVY 235
DB 204 AAAAILAARPKWILGVEGI---GENPICSSTIGHFWGENLEPMDCTP--LKVPANHLL 257
QY 236 SAHDYATSVGQFQWFSPTPPNNMGIWNNKQWGLFNQNIAPVWLGEFGTTL--QSTTDQ 293
DB 258 MPHVGPDVYVQYFNSPDPFPNNMAALWDKHFHFAKAGYA-MAIGFEGKYGEGDPRDI 316
QY 294 TLWKLTVQYLRPTAQYG-ADSFQWTFWSNPDSDGTGGLKDDWQTV--DITVK 343
DB 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGCVWGVNDWTTTPRDDKVK 363

RESULT 15
US-10-369-493-9262
; Sequence 9262, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

167 YTEDESEEDPINTWIEVAKRFKGMVNVIGADLKNPESHVTSPPAAAYTDGTGATWGNPNA 226
176 IDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 226
227 TDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 285
227 LNVP-NRLVSAHDYATSVGPQWFSDDPT-FPNMFGIWNKNGYLFNQNIAPVWLGEFG 284
286 -NLPKNLVSPHYGPDVYNQYFPGAKGPDNLPIWYHHFGVYKLELGYSVVIGEGF 344
285 TTL---QSTTDQWTKLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339
345 GRYGHGDPDRVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 10
US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii

US-10-369-493-20347
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20347
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii

Query Match 38.0%; Score 755; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.3e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDSYMLDQIKSLGYNTIRLPYS 67
Db 54 TSGBE-----TPHLFGVNWGFETPNVHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
QY 68 DILKPGTNPNSINFQMNQDLSQVMDKIVAYAGQIGLRIILDRHRPDCSQSALW 127
Db 108 ESVKFGTQPIGID-YSKNPDRLGDSLQIMEKIKKAGDLGIFVLLDYHRRIGCTHIEPLW 166
QY 128 YTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHD-----PACWGCSDPS 175
Db 167 YTEDESEEDPINTWIEVAKRFKGMVNVIGADLKNPESHVTSPPAAAYTDGTGATWGNPNA 226
QY 176 IDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 226
Db 227 TDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 285
QY 227 LNVP-NRLVSAHDYATSVGPQWFSDDPT-FPNMFGIWNKNGYLFNQNIAPVWLGEFG 284
Db 286 -NLPKNLVSPHYGPDVYNQYFPGAKGPDNLPIWYHHFGVYKLELGYSVVIGEGF 344
QY 285 TTL---QSTTDQWTKLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339
Db 345 GRYGHGDPDRVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 12
US-10-369-493-17534
; Sequence 17534, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17534
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Xylella fastidiosa

US-10-369-493-17534
; Sequence 17534, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17534
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Xylella fastidiosa

Query Match 34.6%; Score 687.5; DB 15; Length 592;

167 YTEDESEEDPINTWIEVAKRFKGMVNVIGADLKNPESHVTSPPAAAYTDGTGATWGNPNA 226
176 IDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 226
227 TDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 285
227 LNVP-NRLVSAHDYATSVGPQWFSDDPT-FPNMFGIWNKNGYLFNQNIAPVWLGEFG 284
286 -NLPKNLVSPHYGPDVYNQYFPGAKGPDNLPIWYHHFGVYKLELGYSVVIGEGF 344
285 TTL---QSTTDQWTKLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339
345 GRYGHGDPDRVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 11
US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii

US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii

US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20347
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii

Query Match 38.0%; Score 755; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.3e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDSYMLDQIKSLGYNTIRLPYS 67
Db 54 TSGBE-----TPHLFGVNWGFETPNVHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
QY 68 DILKPGTNPNSINFQMNQDLSQVMDKIVAYAGQIGLRIILDRHRPDCSQSALW 127
Db 108 ESVKFGTQPIGID-YSKNPDRLGDSLQIMEKIKKAGDLGIFVLLDYHRRIGCTHIEPLW 166
QY 128 YTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHD-----PACWGCSDPS 175
Db 167 YTEDESEEDPINTWIEVAKRFKGMVNVIGADLKNPESHVTSPPAAAYTDGTGATWGNPNA 226
QY 176 IDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 226
Db 227 TDWLAERAGNAVLSPNLLIFVGVQSYN-----GDSYWMGNLQAGQYPPV 285
QY 227 LNVP-NRLVSAHDYATSVGPQWFSDDPT-FPNMFGIWNKNGYLFNQNIAPVWLGEFG 284
Db 286 -NLPKNLVSPHYGPDVYNQYFPGAKGPDNLPIWYHHFGVYKLELGYSVVIGEGF 344
QY 285 TTL---QSTTDQWTKLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339
Db 345 GRYGHGDPDRVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 50
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-09-927-827-50

Query Match 39.9%; Score 793.5; DB 10; Length 535;
Best Local Similarity 46.9%; Pred. No. 1.3e-68;
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVVRVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTIRLP 64
DB 77 YSINSRQIVDSSGKVQLGVNFVGFETGCHVHGLWARKWMDIVQVQGLGFNAVRLP 136

QY 65 YSDIILKPCMTMNSINFYQWMDQGLTSLOYMDKIVAYAGOIGLRILDRHRPDCSGS 124
DB 137 FCPATURSDTMASID-YSRNADLQGLTSLOYMDKIVAYAGOIGLRILDRHRPDCSGS 195

QY 125 ALWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACGCGDPSIDWRLAAR 184
DB 196 ELWYTSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACGCGDPSIDWRLAAR 255

QY 185 AGNAVLSVNPILLIFVEGVQ-----SYNGDSYVWGNLQAGQYPVVNLVNP-NRLVYSAH 238
DB 256 GSAVLAVALAFKLLIAYEGITDNPVCTNG-GIFWGNLQPLACTP--LNIPANRLLAP 312

QY 239 DVATSVGPOTWFSDDPTFPNNPQIKWNGYLFNOMIAPVWLGEFGTTL--QSTTDQWTL 296
DB 313 VYGPVYVOSYFNDSPNNPMAIWEHFQAGTH--ALLGEFGKYGEGDARDKTWQ 370

QY 297 KTLVQYLRPTAQAGDSFQWTFWSNPDSDGTGGILKDDQTV 339
DB 371 DALVKYLR---SKGIN--QGFYWSMNPNSGDTGGILKDDQTV 408

RESULT 8

US-09-888-224-2
; Sequence 2, Application US/09888224
; Patent No. US20020120118A1
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/888,224
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,669
; FILING DATE: 28-Oct-1999
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: 22-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herion
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: Unknown
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-888-224-2

Query Match 38.3%; Score 762; DB 9; Length 553;
Best Local Similarity 45.2%; Pred. No. 1.7e-65;
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNVVRVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTIRLPYSDDLKLP 72
DB 4 VATGEETPIILFVGNWFGFETPNVYVHGLMSRWEDMLLQIKSLGFNAIRLFPCTOSVKP 63

QY 73 GTMENSINFYQWMDQGLTSLOYMDKIVAYAGOIGLRILDRHRPDCSQSALWYTSV 132
DB 64 GTMETAID-YAKNPDQLGLDSVQIMEKLIKAGDLGIFVLDDYHRIGCNFIEPLWYDTSF 122

QY 133 SEATWISDLOALQRYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180
DB 123 SEQDYINTWVEVAQRFKYNWVIGADLKNEPHSSPAPAAAYTDGSGATWGMGNNAIDWNL 182

QY 181 AAEAGNAVLSVNPILLIFVEGVQ-----SYN-GDSYVWGNLQAGQYPVVNLVNP- 230
DB 183 AAEIGRAILLEVAPQWVIFVEGTQFTTPEIDGRYKGNHNAWGNLWGVKYPV--NLPR 240

QY 231 NRLVYSAHAYATSVGPOTWFSDDP--TFPNNPQIKWNGYLFNOMIAPVWLGEFGTTL- 287
DB 241 DKVYVSPQVYGVSEYVQYFP-DPCEGFPDNLPELWYHHPGVKVLDDGYPVVGFGGKYG 299

QY 288 --QSTTDQWTLVQYLRPTAQAGDSFQWTFWSNPDSDGTGGILKDDQTV 339
DB 300 HGGDPRDVTWQNKIIDW--IQNKFCDF--YWSMNPNSGDTGGILKDDQTV 348

RESULT 9

US-10-293-344A-2
; Sequence 2, Application US/10293344A
; Publication No. US20030129723A1
; GENERAL INFORMATION:
; APPLICANT: RAKUTO KASEI INDUSTRIAL CO., LTD. and
; APPLICANT: National Institute of Advanced Industrial Science and Technology
; TITLE OF INVENTION: Thermophilic endoglucanase
; FILE REFERENCE: P017123A
; CURRENT APPLICATION NUMBER: US/10/293,344A
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 2
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
US-10-293-344A-2

Query Match 38.0%; Score 755; DB 14; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.3e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVVRVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTIRLPYSD 67
DB 54 TSGEE-----TPHLFGVNWFGFETPNVYVHGLWARKWMDIVQVQGLGFNAIRLFPCT 107

QY 68 DILKPGTWPNSINFYQWMDQGLTSLOYMDKIVAYAGOIGLRILDRHRPDCSQSALW 127
DB 108 ESKVGTQPIGID-YSKNPDLRGLDSLQIMEKLIKAGDLGIFVLDDYHRIGCNFIEPLW 166

QY 128 YTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
Db 26 YSINNSRQIVDDSGKVQLGAVNVFGETGNHVMHGLWARKWDMIVQMQLGFGNAVRLP 85
QY 65 YSDDILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAVAGQIGLRIILDRHPCDSCGOS 124
Db 86 FCPATLRSDTMPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLLDHHPTDCAGIS 144
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184
Db 145 ELWYTGTSYEAQWLADLRFVANRYKVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
QY 185 AGNAVLSPNLLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVVLNVP-NRLVYSAH 238
Db 205 GSAAVLAVAPKWLIAVEGIDTNPVCSING-GIFWGNLQPLACTP-LNIPANELLAPH 261
QY 239 DYATSVGPQTFWSDPTFPNNMPGINKNNWGYLFNQNIAPVWLGEFGTTL--QSTTDQTLW 296
Db 262 VYGPDPVFQSYFNDSNFPNNMPAIFWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319
QY 297 KTLVOYLPTAQYCADSFQWTFWSNPDSDGTGILKDDWQTV 339
Db 320 DALVKYLR---SKGIN--QGFYWSNPNPNSGDTGILRDDWTSV 357

RESULT 5

US-10-369-493-15740
; Sequence 15740, Application US/10369493
; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 15740

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Xanthomonas campestris

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)...(501)

; OTHER INFORMATION: unsure at all xaa locations

; US-10-369-493-15740

Query Match 39.9%; Score 793.5; DB 15; Length 501;
Best Local Similarity 46.9%; Pred. No. 1.2e-68;
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
Db 26 YSINNSRQIVDDSGKVQLGAVNVFGETGNHVMHGLWARKWDMIVQMQLGFGNAVRLP 85
QY 65 YSDDILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAVAGQIGLRIILDRHPCDSCGOS 124
Db 86 FCPATLRSDTMPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLLDHHPTDCAGIS 144
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184
Db 145 ELWYTGTSYEAQWLADLRFVANRYKVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
QY 185 AGNAVLSPNLLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVVLNVP-NRLVYSAH 238
Db 205 GSAAVLAVAPKWLIAVEGIDTNPVCSING-GIFWGNLQPLACTP-LNIPANELLAPH 261

QY 239 DYATSVGPQTFWSDPTFPNNMPGINKNNWGYLFNQNIAPVWLGEFGTTL--QSTTDQTLW 296
Db 262 VYGPDPVFQSYFNDSNFPNNMPAIFWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319
QY 297 KTLVOYLPTAQYCADSFQWTFWSNPDSDGTGILKDDWQTV 339
Db 320 DALVKYLR---SKGIN--QGFYWSNPNPNSGDTGILRDDWTSV 357

RESULT 6

US-10-369-493-15373
; Sequence 15373, Application US/10369493
; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 15373

; LENGTH: 518

; TYPE: PRT

; ORGANISM: Xanthomonas campestris

; US-10-369-493-15373

Query Match 39.9%; Score 793.5; DB 15; Length 518;
Best Local Similarity 46.9%; Pred. No. 1.3e-68;
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
Db 26 YSINNSRQIVDDSGKVQLGAVNVFGETGNHVMHGLWARKWDMIVQMQLGFGNAVRLP 85
QY 65 YSDDILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAVAGQIGLRIILDRHPCDSCGOS 124
Db 86 FCPATLRSDTMPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLLDHHPTDCAGIS 144
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184
Db 145 ELWYTGTSYEAQWLADLRFVANRYKVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
QY 185 AGNAVLSPNLLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVVLNVP-NRLVYSAH 238
Db 205 GSAAVLAVAPKWLIAVEGIDTNPVCSING-GIFWGNLQPLACTP-LNIPANELLAPH 261
QY 239 DYATSVGPQTFWSDPTFPNNMPGINKNNWGYLFNQNIAPVWLGEFGTTL--QSTTDQTLW 296
Db 262 VYGPDPVFQSYFNDSNFPNNMPAIFWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319
QY 297 KTLVOYLPTAQYCADSFQWTFWSNPDSDGTGILKDDWQTV 339
Db 320 DALVKYLR---SKGIN--QGFYWSNPNPNSGDTGILRDDWTSV 357

RESULT 7

US-09-927-827-50
; Sequence 50, Application US/09927827
; Publication No. US20030036176A1

; GENERAL INFORMATION:

; APPLICANT: Hower, Stanley G.

; APPLICANT: Ramseier, Thomas M.

; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris

; FILE REFERENCE: 38-10(15824)B

; CURRENT APPLICATION NUMBER: US/09/927,827

QY 181 AAFERAGNAVLSVNPILLIFVEGVQSYNGDSYWGNGNLOGAGQYPVVLLNPNRLVYSAHDY 240
Db 222 AAFERAGNAVLSVNPILLIFVEGVQSYNGDSYWGNGNLOGAGQYPVVLLNPNRLVYSAHDY 281
QY 241 ATSVGPQTWFSDDPTFPNNMGPVWKNWGYLFNQNTAPVWLGEFGTTTLOSTTDQWLKTLV 300
Db 282 ATSVGPQTWFSDDPTFPNNMGPVWKNWGYLFNQNTAPVWLGEFGTTTLOSTTDQWLKTLV 341
QY 301 OYLRPTAQYAGDSFQWTFWSPNPSDGTGGILKDDQWQTVDTVKGYLAPIKSSIFDPV 358
Db 342 OYLRPTAQYAGDSFQWTFWSPNPSDGTGGILKDDQWQTVDTVKGYLAPIKSSIFDPV 399

RESULT 2

US-10-360-101-212
; Sequence 212, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of cellulase
US-10-360-101-212

Query Match 52.9%; Score 1052; DB 15; Length 397;
Best Local Similarity 54.5%; Pred. No. 4.4e-94;
Matches 194; Conservative 43; Mismatches 107; Indels 12; Gaps 4;
QY 4 GYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTRL 63
Db 37 GYHTGQNKIVDESSKEAANGNLWFGLETPNTYHLGLWSRMDMDMLDQVKEGYNLRL 96
QY 64 PYSDDLKPGTMBNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQ 123
Db 97 PYSNQLFDSRRPDSID-YHKNPDLVGLNPQIMDKLIEKAGQGIILDRHRPDCSQ 155
QY 124 SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAE 183
Db 156 SELWYTSQYPSERWISDWKMLADRYKNNPTVIGADLHNEPHGQASWGTGNASTDWRLAAQ 215
QY 184 RAGNAVLSVNPILLIFVEGVQ-----SYNGDSYWGNGNLOGAGQYPVVLLNPNRLVYSAHDY 240
Db 216 RAGNAVLSVNPILLIFVEGVQYDHVQNGNSYWGNGNLTGVANYPVVLDPVNRVYSPHDY 275
QY 241 ATSVGPQTWFSDDPTFPNNMGPVWKNWGYLFNQNTAPVWLGEFG--TTLQSTTDQWLKT 298
Db 276 PGVSSQPNWENDAPFNSLPAINDQTWGLYSKQNTAPVVLGEFGGRVNDLSCPEGKQWA 335
QY 299 LVQYLRPTAQYAGDSFQWTFWSPNPSDGTGGILKDDQWQTVDTVKGYLAPIKSS 354
Db 336 LVHYI-----GANNLYFTYWSLNPNSDGTGGILLDDTTTNNRPKQDMLGRIMKPV 385

RESULT 3

US-10-369-493-16379
; Sequence 16379, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16379
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16379

Query Match 40.3%; Score 800.5; DB 15; Length 493;
Best Local Similarity 47.2%; Pred. No. 2.5e-69;
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;
QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTRL 64
Db 26 YSINNSRQIVDDSGKVQLKGVNVFGETGNVHMGILWAERNKDMIVQMGLGFNAVRLP 85
QY 65 YSDDLKPGTMBNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQS 124
Db 86 FCPATLSRDTMPASID-YSRNADLOGLTSLQILDKVIAEFNARGMYVLLDHTPDGAGIS 144
QY 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
Db 145 ELWYTSVTEAQMADLRFAVNRKVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
QY 185 AGNAVLSVNPILLIFVEGVQ-----SYNGDSYWGNGNLOGAGQYPVVLLNPNRLVYSAH 238
Db 205 GSAAVLAVAPKWLIAVEGTTDNPVCSNG-GIFWGGNLQPLACTP--LNPANELLALPH 261
QY 239 DYATSVGPQTWFSDDPTFPNNMGPVWKNWGYLFNQNTAPVWLGEFGTTL--QSTTDQWL 296
Db 262 VYGPDVVFQSYFNDNFNNPALWERHFQFAGTH--ALLGFGGKYGEGDARDXTWQ 319
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSPNPSDGTGGILKDDQWQTV 339
Db 320 DALVKYLR---SKGIN--QGFYWSNPNFSGDTGGILRDDWTSV 357

RESULT 4

US-10-369-493-16123
; Sequence 16123, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16123
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16123

Query Match 39.9%; Score 793.5; DB 15; Length 483;
Best Local Similarity 46.9%; Pred. No. 1.2e-68;
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds
(without alignments)

2716.899 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987

Sequence: 1 AGGGYVHTSGREILDANNVP.....VDFKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1978	99.5	562	9	US-09-981-900B-5
2	1052	52.9	397	15	US-10-360-101-212
3	800.5	40.3	493	15	US-10-369-493-16379
4	793.5	39.9	483	15	US-10-369-493-16123
5	793.5	39.9	501	15	US-10-369-493-15740
6	793.5	39.9	518	15	US-10-369-493-15373
7	793.5	39.9	535	10	US-09-927-827-50
8	762	38.3	553	9	US-09-888-224-2
9	755	38.0	458	14	US-10-293-344A-2
10	755	38.0	458	15	US-10-369-493-1281
11	755	38.0	458	15	US-10-369-493-20347
12	687.5	34.6	592	15	US-10-369-493-17534
13	682.5	34.3	514	15	US-10-369-493-21618
14	660.5	33.2	369	15	US-10-369-493-9473
15	655.5	33.0	421	15	US-10-369-493-9262

16	582	29.3	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	453.5	22.8	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	276.5	13.9	353	15	US-10-369-493-15733	Sequence 15733, A
19	276.5	13.9	353	15	US-10-369-493-16119	Sequence 16119, A
20	276.5	13.9	582	10	US-09-927-827-48	Sequence 48, Appl
21	257.5	13.0	533	12	US-10-424-599-176232	Sequence 176232, A
22	244	12.3	375	15	US-10-369-493-12597	Sequence 12597, A
23	234	11.8	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	191.5	9.6	438	12	US-10-425-114-66753	Sequence 66753, A
25	190.5	9.6	314	15	US-10-369-493-15726	Sequence 15726, A
26	190.5	9.6	314	15	US-10-369-493-16112	Sequence 16112, A
27	190.5	9.6	337	15	US-10-369-493-15359	Sequence 15359, A
28	189.5	9.5	917	16	US-10-437-963-181484	Sequence 181484, A
29	188	9.5	644	16	US-10-437-963-139883	Sequence 139883, A
30	185	9.3	517	12	US-10-107-431-259	Sequence 259, App
31	173.5	8.7	375	10	US-09-917-378-3	Sequence 3, Appli
32	173.5	8.7	375	10	US-09-917-378-6	Sequence 6, Appli
33	173.5	8.7	762	10	US-09-917-378-1	Sequence 1, Appli
34	168.5	8.5	346	10	US-09-917-378-8	Sequence 8, Appli
35	168	8.5	356	10	US-09-917-378-7	Sequence 7, Appli
36	166.5	8.4	558	16	US-10-437-963-185049	Sequence 185049, A
37	160.5	8.1	722	12	US-10-282-122A-51993	Sequence 51993, A
38	156.5	7.9	518	10	US-09-769-734-56	Sequence 56, Appl
39	155.5	7.8	356	15	US-10-369-493-17756	Sequence 17756, A
40	153.5	7.7	1449	16	US-10-437-963-176100	Sequence 176100, A
41	152.5	7.7	533	12	US-10-424-599-150988	Sequence 150988, A
42	150.5	7.6	268	12	US-10-424-599-284543	Sequence 284543, A
43	150.5	7.6	315	15	US-10-369-493-9177	Sequence 9177, Ap
44	147	7.4	217	16	US-10-437-963-112718	Sequence 112718, A
45	146	7.3	313	15	US-10-369-493-15674	Sequence 15674, A

ALIGNMENTS

RESULT 1

US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US20020138878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Macboon, Shatina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE TO FERMENTABLE SUGARS
; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-639
; CURRENT APPLICATION NUMBER: US/09/981,900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR FILING DATE: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5

Query Match	99.5%	Score	1978	DB	9	Length	562
Best Local Similarity	99.7%	Pred. No.	1.6e-184				
Matches	357	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
QY	1	AGGGYVHTSGREILDANNVPVRIAGINWFGPFCNYYVHGLWSRDYRSMLEQIKSLGNT	60				
Db	42	AGGGYVHTSGREILDANNVPVRIAGINWFGPFCNYYVHGLWSRDYRSMLEQIKSLGNT	101				
QY	61	IRLPYSDDLKPGMNSINFYQVNDLQGLTSLQVMDKIVAYAGQIGLRILDRHPDC	120				
Db	102	IRLPYSDDLKPGMNSINFYQVNDLQGLTSLQVMDKIVAYAGQIGLRILDRHPDC	161				
QY	121	SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	180				
Db	162	SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	221				

US-08-525-697-2
; Sequence 2, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Helst-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-525-697-2
Query Match 8.5%; Score 168.5; DB 1; Length 377;
Best Local Similarity 23.1%; Pred. No. 5.4e-08;
Matches 82; Conservative 56; Mismatches 116; Indels 101; Gaps 21;
Qy 24 AGIN--WFGETCNVYVHGLWSRYSMLDQIKSLGYNTIRLPYSDDLKPGTWPNSINF 81
Db 50 AGTNSYIGFETLNDDVDLWSQLAASDLKILRVWGFNDVTKPTD-----GTV-----W 99
Qy 82 YQNMCD-----LQGLTSLQVMDKIVAYAGQIGLRILD--RHRPDCSGQSAL----- 126
Db 100 YQLHANGTSTINTGADGLRLDYVTVTSAEKYGVKLIINFVNEWTDYGMQAYVTAYGAAA 159
Qy 127 ---WYTSVSEATWISDLQALAQRYKNPTVGFDLHNEPHDPCWCGGSPSIDWRLAAE 193
Db 160 QTDFTYNTAQAYKNIKAVVSYSSAIFAWELEANEPR---CQGC-DTSVLYNWISD 215
Qy 184 RAGNAVLNVPNLLIFVEGVQSVNGDSYWMGNNLQAGQYPPV---LNVNRLVYSADH 239
Db 216 -TSKYIKSLQSKLVTI-----GDE-GLGLVDSDGSYPYTYGEGLNFTKNLGISTID 266
Qy 240 YAT-SVGPQTWFSPTFPNNMGIWKNWGLVFNQIA-----PWLGEFGTT----- 286
Db 267 FGLTLHPDWSMGTSYD-----WGNW---ITAHAAACAVKCKPCLLEEGYGVGNHCA 315
Qy 287 LQSTTDDT-----WLKTLVQLRPTAQYGADSFQWTF-WSWNPDSGDT 328
Db 316 VESPWQQTAGNATGISGLYN-----QYGT-----TFSWQSPNDGNT 353

RESULT 13
US-09-797-464A-4
; Sequence 4, Application US/09797464A
; Patent No. 6630340
; GENERAL INFORMATION:
; APPLICANT: Wiltting, Reinhard
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: Family 5 Xyloglucanases
; FILE REFERENCE: 6073.200-US
; CURRENT APPLICATION NUMBER: US/09/797,464A
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Paenibacillus sp.
US-09-797-464A-4
Query Match 7.2%; Score 143; DB 4; Length 400;
Best Local Similarity 22.5%; Pred. No. 1.8e-05;
Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;
Qy 3 GGYWHTSGREILDANNVPRVRIAGINWFGFETCNVYVHGLWSRYSMLDQIKSLGYNTIR 62
Db 48 GAGWNLGNQLBATVNGVPSETA--W-----GNPVV-----TPELIKKYKAAAGFKTIR 92
Qy 63 LP--YSDDLKPGTWPN--SINFYQNNQDLQGLTSLQVMDKIVAYAGQIGLRILDHREPD 119
Db 93 IPVSYLNIH--GSAPNTYTIINAAMLR-----VQTVDVAYNEGLVYVINIHGDG 139
Qy 120 CSGQSALWYTSVSEATWISD---LQALAQRVKNPTVVGFDLHNEPHDPCWCGGCDP 174
Db 140 YNSIPGGWLLVNGSNQAAIKEKYQKVWQIATKSNYNERLIFESMNEVFD---GNYGNP 196
Qy 175 SIDWELAAERAGNAVLNVPNLL--IFVEGVQSV--NODSYW-----WGNLQ--GAGQYP 224
Db 197 -----NAAYYANLNAYNQIFVDYTVRTGNNARWLLIPGWNTNIDYTVNGY 244
Qy 225 VVL-----NVPN---RLVYSADY-----ATSVGP---QTFWSDP 253
Db 245 PALPTDHFRRSAIPSSQKRIMISAHYSPWDFPAGEENGNIQWGNAAITNPKSKTWGQED 304
Qy 254 TFPNNMPCIMWK--NWGYLFNQNIAPVWLGBFGT---TLQSTTDTQTLVQLVLRPTA- 307
Db 305 YLNAQCFKSMYDKFVTQGY-----PVVIGFSGIDKTAIDSTNNVYRQAYAKAVTATAK 357
Qy 308 QYGADSFQWTFWSWNPDSGDTG 329
Db 358 KYGAVPVTW-----DMGNHG 372
RESULT 14
US-09-797-464A-7
; Sequence 7, Application US/09797464A
; Patent No. 6630340
; GENERAL INFORMATION:
; APPLICANT: Wiltting, Reinhard
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: Family 5 Xyloglucanases
; FILE REFERENCE: 6073.200-US
; CURRENT APPLICATION NUMBER: US/09/797,464A
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 348
; TYPE: PRT

QY 338 TVDTVKDGYLAP 349
 Db 562 TWDEQKYNFLKP 573

RESULT 10

US-09-136-574A-43

; Sequence 43, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 1426 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e

SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 31.3%; Score 621; DB 3; Length 1426;

Best Local Similarity 35.5%; Pred. No. 4.8e-51;

Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNPVRIAGINWFGFTCNVYVHGLSRDYSRMLDQIKSLGYNVIRLPYSDDI 69

Db 1022 GNKIVDKGDFVWLTVGNVFGFTGTVNFGVNSCNLSALAIANRGNLVRPISAEI 1081

QY 70 L---KPTMPN-SINFQMNQDQLGTLQSLQMDKIVAYACQIGLRILDRH--RPCSGQ 123

Db 1082 ILNWSKGIYKPKNINYY-VNPEGLGTLSEVDFVVKTCVEGLKIMLDTHSAKTAMGH 1140

QY 124 -SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEHPD-----ACWGCDDPS 175

Db 1141 IYPWVYTDITPDIYKACEWITERYKNDDTIVAFDLKNEPHKPKWDVSFAKWDNSTDI 1200

QY 176 IDWELAAERAGNAVLISVNPILLIFVSGVQSYNGD-----SYWGGNLOGAGQ 222
 Db 1201 NNKYYAAETCAKRLAKNPNLLIVIEGIEAYPKDDVTWTSKSSDDYYSIWGGNLRGVKK 1260
 QY 223 YPVVL-NVPNRLVYSAHDYATSYGPTW-----FSDPTFPNMPGINKWNGYLFNQNIAP 277
 Db 1261 YPINLGOYQNKVYSPHDYGLVYQQPWFYFGFTKDTLYND---CWRDNWYIIMDNGIAP 1317
 QY 278 VWLGEFGTTLQSTTDQTLWKLTVQYLRPTAQYGADSFQMTFWSNPDSPGTGGILKDDWQ 337
 Db 1318 LLIGEWGGYLDGGDNKWKMTYLRDYII-----ENIHHTFWCYNANSBDTGLGVGYDFS 1371
 QY 338 TVDTVKDGYLAP 349
 Db 1372 TWDEQKYNFLKP 1383

RESULT 11

US-08-276-213-2

; Sequence 2, Application US/08276213

; Patent No. 5536655

; GENERAL INFORMATION:

APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert

APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: USA

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,213

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

US-08-276-213-2

Query Match 10.3%; Score 205; DB 1; Length 38;

Best Local Similarity 97.4%; Pred. No. 4.9e-13;

Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGWHTSGREILDANNVPVRIAGINWFGFTCNVYV 38

Db 1 AGGGWHTSGREILDANNVPVRIAGINWFGFTCNVYV 38

RESULT 12

Db 199 TDMNLAERIGKALVKAPHLIFVEGTQFTNPKTSSYKXGYNAMWGNLMAVKDYPV- 257
Qy 227 LNVP-NRLVYSAHDYATSVGPQTWFSDEPT-FPNMYPGIMNKNWGYLNFQNIAPVWILGEFG 284
Db 258 -NLRNKLIVSPHYGPDVYNQPYFGPAKGPPDNLPDIWYHFGYKLELGYSVWIGFEG 316
Qy 285 TTL---QSTTDQWLKTLVQYLRPTAQYGADSPQWTFWNSNPDGDTGGILKDDQWTV 339
Db 317 GYGHGGPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 369

RESULT 8

US-09-869-197-5
; Sequence 5, Application US/09869197
; Patent No. 6566113
; GENERAL INFORMATION:
; APPLICANT: TAKAYAMA, Masanori
; APPLICANT: UMEDA, Kahoko
; APPLICANT: KAYAMA, No. 6566113uto
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunosai
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: TAKAYAMA-6
; CURRENT APPLICATION NUMBER: US/09/869,197
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/JP99/07009
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: JP 366237/1998
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii OT3
US-09-869-197-5

Query Match 38.0%; Score 755; DB 4; Length 458;
Best Local Similarity 45.0%; Pred. No. 8.5e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;
Qy 8 TSGREILDANNVPRIAGINFGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLPYSDDI 67
Db 54 TSGBE-----TPIHLFGVNWFGFETPNVHVHGLWKENWEDMLLIQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTNPNSINFYQMNODLQGLTSLQVMDKIVAVAGOIGLRIILDRHPDCSGOSALW 127
Db 108 ESKVPGTQPID-YKSNPDRLGDSLQIKWEKIKKAGDLGIFVLLDYHRIIGCTHIEPLW 166
Qy 128 YTSVSEATWISDLQALAQRYKGNPTVYVGFDLNNEPHD-----PACWCGCDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRFGKYNVIGADLXNEPHSVTSPPAAVTDGTGATWGMGNPA 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYWGNNLQAGQYPPV 226
Db 227 TDMNLAERIGKALVKAPHLIFVEGTQFTNPKTSSYKXGYNAMWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVGPQTWFSDEPT-FPNMYPGIMNKNWGYLNFQNIAPVWILGEFG 284
Db 286 -NLRNKLIVSPHYGPDVYNQPYFGPAKGPPDNLPDIWYHFGYKLELGYSVWIGFEG 344
Qy 285 TTL---QSTTDQWLKTLVQYLRPTAQYGADSPQWTFWNSNPDGDTGGILKDDQWTV 339
Db 345 GYGHGGPRDVIWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 9

US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 31.3%; Score 621; DB 3; Length 616;
Best Local Similarity 35.5%; Pred. No. 1.3e-51;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;
Qy 10 GREILDANNVPRIAGINFGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLPYSDDI 69
Db 212 GNKIVDKGKPVLTGVNWFNTGTVNFDGVWSCNLKALAEIANRGNFNLRLVPISAEL 271
Qy 70 L---KEGTMEN-SINFYQMNODLQGLTSLQVMDKIVAVAGOIGLRIILDRH--RPDCSGQ 123
Db 272 ILNWKGIYKPNINYY-VNPELEGLTSLVDFVFKVTKCVGLKIMLDIHSKTDAMGH 330
Qy 124 -SALWYTSVSEATWISDLQALAQRYKGNPTVYVGFDLNNEPHD-----ACWCGCDPS 175
Db 331 IYVWYTTITIPEDYKACEWITERYKNDTIVAFDLKNEPHGKPPQWDSVFAKWDNSTDI 390
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----SYWNGNLOAGQ 222
Db 391 NNWYAAETCAKRIILAKPNMLIVIEGLEATPKDDVTWTSKSSDYSTWNGNLRGVKK 450
Qy 223 YPVVL-NVPENLVYSAHDYATSVGPQTW----FSDPTFPNNMYPGIMNKNWGYLNFQNIAP 277
Db 451 YPINLGQYQNKVYSPHYGPDVYNQPYFGPFQTKDTLYND---CWRDNWYIMDNGIAP 507
Qy 278 VWLGEFGTTLQSTTDQWLKTLVQYLRPTAQYGADSPQWTFWNSNPDGDTGGILKDDQW 337
Db 508 LLIGEWGGLDGGDNKRWNTYLRDYYI-----ENHIIHFWCYNANSGLDGLGVGYDFS 561

Query Match 99.5%; Score 1978; DB 1; Length 358;
Best Local Similarity 99.7%; Pred. No. 1.1e-183;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120

QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180
DB 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPPVVLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPPVVLNPNRLVYSAHDY 240

QY 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNGYLFNONTAPVNLGEGFTTLOSTTDOQWLKTLV 300
DB 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNGYLFNONTAPVNLGEGFTTLOSTTDOQWLKTLV 300

QY 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2
US-08-276-213-3
; Sequence 3, Application US/08276213
; Patent No. 5536655
; GENERAL INFORMATION:
; APPLICANT: Thomas, Steven
; APPLICANT: Laymon, Robert
; APPLICANT: Himmel, Michael
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,213
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08
; TELEPHONE: (303)231-1000
; TELEFAX: (303)231-1098
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-276-213-3

Query Match 99.5%; Score 1978; DB 1; Length 521;
Best Local Similarity 99.7%; Pred. No. 2e-183;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120
DB 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120

QY 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180
DB 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPPVVLNPNRLVYSAHDY 240
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNQLQAGQYPPVVLNPNRLVYSAHDY 240

QY 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNGYLFNONTAPVNLGEGFTTLOSTTDOQWLKTLV 300
DB 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNGYLFNONTAPVNLGEGFTTLOSTTDOQWLKTLV 300

QY 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358
DB 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3
US-08-651-572-2
; Sequence 2, Application US/08651572
; Patent No. 5789228
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CECCHIA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Heiron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-651-572-2

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds
(without alignments)
1260.143 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987

Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDFPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1978	99.5	358	1	US-08-604-913B-11
2	1978	99.5	521	1	US-08-276-213-3
3	762	38.3	553	3	US-08-651-572-2
4	762	38.3	553	3	US-09-066-544-2
5	762	38.3	553	3	US-08-951-086-2
6	762	38.3	553	3	US-09-430-669-2
7	755	38.0	430	4	US-08-869-197-1
8	755	38.0	458	4	US-09-869-197-5
9	621	31.3	616	3	US-08-136-574A-47
10	621	31.3	1426	3	US-08-136-574A-43
11	205	10.3	38	1	US-08-276-213-2
12	168.5	8.5	377	1	US-08-525-697-2
13	143	7.2	400	4	US-08-797-464A-4
14	139	7.0	348	4	US-08-797-464A-7
15	139	7.0	476	4	US-09-339-159B-4
16	139	7.0	490	4	US-08-339-159B-2
17	136	6.8	395	4	US-09-797-464A-2
18	136	6.8	363	4	US-08-485-648-4
19	136	6.8	468	4	US-09-503-565-4
20	136	6.8	468	4	US-09-485-649-4
21	136	6.8	468	4	US-09-339-159B-8
22	136	6.8	493	4	US-09-485-648-2
23	136	6.8	493	4	US-09-503-565-2
24	136	6.8	493	4	US-09-485-649-2
25	136	6.8	493	4	US-09-339-159B-6
26	136	6.8	493	4	US-08-276-213-1
27	131	6.6	24	1	US-08-276-213-1

ALIGNMENTS

RESULT 1

US-08-604-913B-11
; Sequence 11, Application US/08604913B

; Patent No. 5712142

; GENERAL INFORMATION:

; APPLICANT: Adney, William S.

; APPLICANT: Thomas, Steven R.

; APPLICANT: Himmel, Michael E.

; APPLICANT: Baker, John O.

; APPLICANT: Chou, Yat-Chen

; TITLE OF INVENTION: METHOD FOR INCREASING

; THERMOSTABILITY IN CELLULASE ENZYMES

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBC PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASC II (DOS) text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/604,913B

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/276,213

; FILING DATE: 15-07-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Edna M. O'Connor

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: 95-56

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/384-7573

; TELEFAX: 303/384-7499

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 358 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

; FEATURE:

; NAME/KEY: E1-CAT

; US-08-604-913B-11

Sequence 12, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 22, Appl
Sequence 61, Appl
Sequence 25, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 27, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 16, Appl
Sequence 16, Appl

Db 183 AAEIRGRAILEVAPQWIFVEGTFPTPEIDGRYKGNHAWGGNLMGVKRPV--NLPR 240
QY 231 NRVLYSAHDYATSGVQTWFSDP--TFPNNMPCIMNKNWGYLFNQNIAPVWLGEFGTTL- 287
Db 241 DKLIVSPQVYGGPDVQPYF-DPEGFPDNLPEIWHYHFGYVKLDLGYPPVIGFGRKYG 299
QY 288 --QSTTDQTLVLYLRTAQYAGDSFQWTFWSNPDGSGTGILKDDWQTV 339
Db 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YMSWNPNSGDTGGILKDDWTTI 348

RESULT 10
AAW34985
ID AAW34985 standard; protein; 841 AA.
XX AC AAW34985;
XX 27-AUG-2003 (revised)
DT 21-MAY-1998 (first entry)
XX DE Archaeobacterial thermostable endoglucanase.
XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW archaeal bacterium; thermostable enzyme; thermophilic.
XX Unidentified.
XX WO9744361-A1.
XX 27-NOV-1997.
XX 22-MAY-1997; 97WO-US008793.
XX 22-MAY-1996; 96US-00651572.
XX (RECO-) RECOMBINANT BIOTOCALYSIS INC.
XX Lam DE, Mathur EJ;
XX WPI; 1998-018435/02.
XX N-PSDB; AAT94193.
XX Endo:glucanase(s), preferably from archaeal bacterium, AEPII 1a - useful
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
PT bonds in cellulose.
XX Claim 1; Fig 1A; 164pp; English.
XX This protein comprises a thermostable 60.9 kDa endoglucanase of archaeal
CC bacterial strain AEPII1a, a thermophilic isolate of a marine hydrothermal
CC vent. The endoglucanase is capable of degrading carboxymethylcellulose
CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It can be
CC produced from native cells or from recombinant host cells, especially
CC prokaryotic host cells transformed with a plasmid or virus-derived vector
CC including the endoglucanase DNA (see AAT94193). 23 Other, homologous
CC endoglucanases (see AAW34986-W35008) are also claimed. The endoglucanase
CC can be used to degrade cellulose for the conversion of plant biomass into
CC fuels and chemicals; for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices. (Updated on 27-AUG-2003 to
XX correct OS field.)
XX Sequence 841 AA;
Query Match 38.4%; Score 763; DB 2; Length 841;
Best Local Similarity 45.2%; Pred. No. 4.9e-63;
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;
QY 13 ILDANNVPRVIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYTILPYSDDLKP 72

Db 4 VATGETPIHLFGVNWFGFETPNVYVHGLMSRNWEDMLQIKSLGFNAIRLPFTQSVKP 63
QY 73 GTMENSINFCMQODLGLTSLQVMDXIVAVAGIIGRIIILDRHRPDCSGQSALWTSSV 132
Db 64 GTMPTAID-YAKNPDLQGLDSVQIMEXIIKAGDLGIFVILDRHRCNIEPLWYDTSF 122
QY 133 SEATWISDLOALAQRYKGNPTVVGFDLNEPHDP-----ACWCGDPSIDWRL 180
Db 123 SEQDYINTWVEVAQRFCKYNNVIGADLNKNEPHSSPAPAAAYTDGSGATWGNNAIDWNL 182
QY 181 AAEIRAGNAVLSVNPILLIFVEGVQ-----SYN-CDSYWAGNLOGAGQYPVLVNP- 230
Db 183 AAEIRGRAILEVAPQWIFVEGTFPTPEIDGRYKGNHAWGGNLMGVKRPV--NLPR 240
QY 231 NRVLYSAHDYATSGVQTWFSDP--TFPNNMPCIMNKNWGYLFNQNIAPVWLGEFGTTL- 287
Db 241 DKLIVSPQVYGGPDVQPYF-DPEGFPDNLPEIWHYHFGYVKLDLGYPPVIGFGRKYG 299
QY 288 --QSTTDQTLVLYLRTAQYAGDSFQWTFWSNPDGSGTGILKDDWQTV 339
Db 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YMSWNPNSGDTGGILKDDWTTI 348

RESULT 11
ABG70759
ID ABG70759 standard; protein; 553 AA.
XX AC ABG70759;
XX 09-DEC-2002 (first entry)
XX DE T. maritima endoglucanase protein.
XX Endoglucanase; enzyme; cellulose; plant cell wall; homopolysaccharide;
KW D-glucose; beta configuration; beta 1->4 glycosidic bond; hydrolysis;
KW 1,4-beta-D-glucan glucanohydrolase; exoglucanase;
KW 1,4-beta-D-glucan cellobiohydrolase; thermostable; catalysis; antigen.
XX Thermotoga maritima.
XX US2002120118-A1.
XX 29-AUG-2002.
XX 22-JUN-2001; 2001US-00888224.
XX 22-MAY-1996; 96US-00651572.
XX 24-APR-1998; 98US-00065544.
XX 28-OCT-1999; 99US-00430669.
XX (SHOR/) SHORT J M.
XX (LAMD/) LAM D E.
XX (MATH/) MATHUR E J.
XX Short JW, Lam DE, Mathur EJ;
XX WPI; 2002-691213/74.
XX N-PSDB; ABS54439.
XX New isolated nucleic acid encoding a polypeptide having endoglucanase
PT activity, useful as a probe for isolating or identifying other
PT endoglucanase genes having a sequence that is similar to its sequence.
XX Claim 35; Fig 5; 36pp; English.
XX The invention discloses an isolated nucleic acid encoding a polypeptide
CC from Thermotoga maritima which has an endoglucanase activity. Cellulose
CC is a tough, fibrous, water insoluble substance found in the cell walls of
CC plants. It consists of an unbranched homopolysaccharide of 10,000 to
CC 15,000 D-glucose units in a beta configuration linked by beta 1->4
CC glycosidic bonds. The enzymatic hydrolysis of cellulose requires the
CC action of both endoglucanases (1,4-beta-D-glucan glucanohydrolase) and
CC exoglucanases (1,4-beta-D-glucan cellobiohydrolase). The endoglucanase of

KW emulsion stability; temperature resistance; pseudoplasticity; amylase;
 KW cellulase; extracellular protease; intracellular protease;
 KW glucose dehydrogenase; enzyme.
 OS Xanthomonas campestris.
 XX
 PN US2003036176-A1.
 XX
 PD 20-FEB-2003.
 XX
 XX 10-AUG-2001; 2001US-00927827.
 XX
 PR 28-MAR-2001; 2001US-0279493P.
 XX
 XX (BOWE/) BOWER S G.
 PA (RAMS/) RAMSEIER T M.
 XX
 XX Bower SG, Ramseier TM;
 XX WPI; 2003-625389/59.
 DR N-PSDB; ADD24896.
 XX
 PT New transformed cell or organism having reduced or enhanced activity of
 PT at least one protein, useful for producing xanthan gum, which are useful
 PT for providing formulations and properties.
 XX
 XX Claim 1; SEQ ID NO 50; 135pp; English.
 XX
 CC The present invention relates to polypeptide and polynucleotide sequences
 CC from Xanthomonas campestris which may be used for activity reduction or
 CC enhancement using directed genetic engineering. A transformed cell or
 CC organism having reduced or enhanced activity of at least one such protein
 CC e.g. galactomannanase can be generated by disrupting the gene encoding
 CC the protein. The activity of the protein is reduced by the presence of an
 CC antisense nucleic acid sequence. The nucleic acid sequence of the gene
 CC encoding the protein is a recombinant sequence having at least one
 CC mutation as compared to the wild-type gene encoding the protein. The
 CC transgenic cell or microorganism are useful for producing xanthan gum,
 CC which are useful for providing formulations and properties, such as long-
 CC term suspension and emulsion stability in alkaline, acid, and salt
 CC solutions, temperature resistance, and pseudoplasticity. The present
 CC sequence represents an enzyme relating to the present invention.
 XX
 SQ Sequence 535 AA;

Query Match 39.9%; Score 793.5; DB 7; Length 535;
 Best Local Similarity 46.9%; Pred. No. 3.2e-66;
 Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;
 QY 5 YNHTGREILDANNVPRIAGINWFGFETCNVYVHGLMSRDYRSMDDQIKSLGYNTIRLP 64
 DB 77 YSINNSRQIVDDSGKVQLKGVNFGFETGNHVMHGLMARNMKDMIVQVQGLGFNAVRLP 136
 QY 65 YSDDILKPCITMPSINFYQMNODLOGLTSLQVMDKIVAVAGQIGLRILDRHRPDCSQS 124
 DB 137 FCPATLRSTMPASID-YSRNADLOGLTSLQLDKIVAEFNAGMYVLLDHTTPCAGIS 195
 QY 125 ALWYTSVSEATWISDLQALQRYKGNPTVTFGLHNEPHDPACWCGDPSIDWFLAAER 184
 DB 196 ELWYTGSYTEAQWLADLRVANKYKNVPVVLGLDLKNEPHGAATWGTGNAATDNWKAER 255
 QY 185 AGNAVLSVNPILLIFVEGVQ-----SYNGDSYVWCGNLCQAGQYVVLNVP-NRLVYSAH 238
 DB 256 GSAALVAVAPKWLIAVEGT-TDNPVCTNG-GIFWCGNLCPLACTP--LNIPANRLLALPH 312
 QY 239 DYATSVGPOTWFSDDPTFPNNMPCGNKWNKGVFNQNIAPVWLGEFTGL--QSTTDDQWL 296
 DB 313 VYGPDPVYQSYFNSFPNMPAIWERHFGQFAGTH--ALLGEFDGKYGEGDARDTKTQ 370
 QY 297 KTLVQLRPTAQYADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
 DB 371.DALVYLR---SKGIN--QGFYWSNPNDSGDTGGILRDDWTSV 408

RESULT 9
 AAW34999
 ID AAW34999 standard; protein; 628 AA.
 XX
 AC AAW34999;
 XX
 DT 27-AUG-2003 (revised)
 DT 21-MAY-1998 (first entry)
 XX
 DE Archaeobacterium AEP11a endoglucanase.
 XX
 KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
 KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
 KW thermostable enzyme; thermophilic; glycosidase.
 XX
 OS archaeson.
 OS Unidentified.
 XX
 FN WO9744361-A1.
 XX
 PD 27-NOV-1997.
 XX
 PF 22-MAY-1997; 97WO-US008793.
 XX
 PR 22-MAY-1996; 96US-00651572.
 XX
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 XX
 PI Lam DE, Mathur EJ;
 XX
 DR WPI; 1998-018435/02.
 DR N-PSDB; AAT94207.
 XX
 PT Endo:glucanase(s), preferably form archaeal bacterium, AEP11a - useful
 PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
 PT bonds in cellulose.
 XX
 PS Claim 1; Fig 10; 164pp; English.
 XX
 CC This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone
 CC 63GP2), a hydrothermal vent isolate. The endoglucanase is capable of
 CC degrading carboxymethylcellulose and of hydrolysing the beta-1,4-
 CC glycosidic bonds in cellulose. It has homology to another endoglucanase
 CC (see AAW34985) of archaeobacterium AEP11a. It can be produced from native
 CC cells or from recombinant host cells, especially prokaryotic host cells
 CC transformed with a plasmid or virus-derived vector including the
 CC endoglucanase DNA (see AAT94207). 24 endoglucanases (see AAW34986-W35008)
 CC are claimed. They can be used to degrade cellulose for the conversion of
 CC plant biomass into fuels and chemicals, for use in detergents, textiles,
 CC animal feed, waste treatment, and in the fruit juice and brewing
 CC industries for the clarification and extraction of juices. (Updated on 27
 CC -AUG-2003 to correct OS field.)
 XX
 SQ Sequence 628 AA;
 Query Match 38.4%; Score 763; DB 2; Length 628;
 Best Local Similarity 45.5%; Pred. No. 3.2e-63;
 Matches 161; Conservative 48; Mismatches 109; Indels 36; Gaps 11;
 QY 13 ILDANNVPRIAGINWFGFETCNVYVHGLMSRDYRSMDDQIKSLGYNTIRLPYSDILXP 72
 DB 4 VATEGTEPIHLFGVNWFGFETPNVYVHGLMSRWNEMLQIKSLGNALRLPCTQSVKP 63
 QY 73 GTWPNFINFYQMNQDQGLTSLQVMDKIVAVAGQIGLRILDRHRPDCSQSALWYTSV 132
 DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEKIIKKGADLGIFVLDDYHRIKCNFIEPLWYDTSF 122
 QY 133 SEATWISDLQALQRYKGNPTVTFGLHNEPHDP-----ACWCGDPSIDWFL 180
 DB 123 SEQDYINTWVEVAQRFGKYWNVIGADLRKNEPHSSSPAPAYTDCSGATWGMGNATDNL 182
 QY 181 AABRAGNAVLSVNPILLIFVEGVQ-----SYN-GDSYVWCGNLCQAGQYVVLNVP- 230

CC activity comprises replacing an active site-associated amino glycosyl-
CC stabilising amino acid with an amino acid that does not strongly bind a
CC disaccharide product in the active site. Conversely, the method for
CC making a glycosyl hydrolase with increased soluble substrate catalytic
CC activity comprises replacing a hydrophobic substrate-binding amino acid
CC with a positively charged residue. The invention also discloses mutants
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
CC produced according to the method of the invention. The Y245G mutant
CC (AAB48786) has improved activity with insoluble substrates, and the W42R
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase Y82R mutant
XX
SQ Sequence 521 AA;
Query Match 99.1%; Score 1969; DB 4; Length 521;
Best Local Similarity 99.4%; Pred. No. 7.2e-178;
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT 60
DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT 60
QY 61 IRLPYSDILKPGTNPNSINFYQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDILKPGTNPNSINFYQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAB48786 standard; protein; 521 AA.
DB 181 AAB48786;
XX AC AAB48786;
XX DT 09-MAR-2001 (first entry)
XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.
XX KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
XX OS cellulose hydrolysis; ethanol production; fermentation; mutant; muten.
XX OS Acidothermus cellulolyticus.
XX OS Synthetic.
XX FN WO200070031-A1.
XX PD 23-NOV-2000.
XX PF 19-MAY-2000; 2000WO-US013971.
XX PR 19-MAY-1999; 99US-0134925P.
XX PA (MIDE) MIDWEST RES INST.
XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
PI Decker SR;
XX WPI; 2001-061226/07.
XX Preparation of glycosyl hydrolase with an increased catalytic activity on
PT insoluble substrate.
XX Claim 16; Page 22-24; 30pp; English.
XX The invention relates to a method for making glycosyl hydrolase mutants
CC with increased catalytic activity with either insoluble or soluble
CC cellulose substrates relative to the wild-type enzyme. The method for
CC making a glycosyl hydrolase with increased insoluble substrate catalytic
CC activity comprises replacing an active site-associated amino glycosyl-
CC stabilising amino acid with an amino acid that does not strongly bind a
CC disaccharide product in the active site. Conversely, the method for
CC making a glycosyl hydrolase with increased soluble substrate catalytic
CC activity comprises replacing a hydrophobic substrate-binding amino acid
CC with a positively charged residue. The invention also discloses mutants
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)
CC produced according to the method of the invention. The Y245G mutant
CC (AAB48786) has improved activity with insoluble substrates, and the W42R
CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with
CC soluble substrates. The invention also encompasses DNA encoding these
CC mutants. The glycosyl hydrolases of the invention are used as catalysts
CC for cellulose hydrolysis to produce sugars that can be fermented to
CC produce fuels such as ethanol. The present sequence represents the
CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant
XX
SQ Sequence 521 AA;
Query Match 98.8%; Score 1964; DB 4; Length 521;
Best Local Similarity 99.4%; Pred. No. 2.2e-177;
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT 60
DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLDQIKSLGYNT 60
QY 61 IRLPYSDILKPGTNPNSINFYQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
DB 61 IRLPYSDILKPGTNPNSINFYQMNODLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAB48786 standard; protein; 521 AA.
DB 181 AAB48786;
XX AC AAB48786;
XX DT 09-MAR-2001 (first entry)
XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.
XX KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;
XX OS cellulose hydrolysis; ethanol production; fermentation; mutant; muten.
XX OS Acidothermus cellulolyticus.
XX OS Synthetic.
XX FN WO200070031-A1.
XX PD 23-NOV-2000.
XX PF 19-MAY-2000; 2000WO-US013971.
XX PR 19-MAY-1999; 99US-0134925P.
XX PA (MIDE) MIDWEST RES INST.
XX

Db 282 ATSVYPTQWFSDFPTFPNNPFGIWNKNGVLFNQNIAPIVNLGEGFTTLOSTTDQIWLKTLV 341

Qy 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358
 Db 342 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 399

RESULT 5

AAI69508
 ID AAY69508 standard; protein; 562 AA.

AC AAY69508;
 DT 10-APR-2000 (first entry)

XX Acidothermus cellulolyticus E1 endoglucanase.
 XX E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;
 KW beta-1,4-endoglucanase; endocellulase; thermostable.
 XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers
 FT Peptide 1..41 /note= "Putative signal peptide"
 FT Peptide 14..41 /note= "Putative signal peptide (alternative)"
 FT Domain 42..404 /note= "Catalytic domain"
 FT Region 405..460 /note= "Linker region"
 FT Domain 461..562 /note= "Cellulose binding domain (CBD)"

XX CA2226898-A1.
 XX 25-SEP-1999.
 XX 25-MAR-1998; 98CA-02226898.
 XX 25-MAR-1998; 98CA-02226898.
 XX (MIDE) MIDWEST RES INST.

XX Laymon RA, Adney WS, Thomas SR, Himmel ME;
 XX WPI; 2000-087663/08.
 XX N-PSDB; AA255924.

XX Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful
 PT for labeling or modifying a cellulose and for purifying or immobilizing a
 PT binding domain fusion protein to cellulose.

XX Claim 1; Fig 2; 85pp; English.
 XX This sequence represents the Acidothermus cellulolyticus E1
 CC endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The
 CC cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which
 CC encode it are specifically claimed. The CBD is believed to be roughly
 CC wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge
 CC tip is inserted between the microfibrils of the cellulose fibre,
 CC disrupting the crystalline structure, and making the cellulose linkages
 CC more accessible to the catalytic domain of the E1 endoglucanase. The E1
 CC endoglucanase CBD is useful in labelling or modifying the surface of
 CC cellulose or other polysaccharides. Such modified cellulose can then be
 CC used in textile, pulp, paper, chemical and pharmaceutical industries.
 CC CBDs can be used in affinity purification of CBD-fusion proteins, and can
 CC also be used to immobilise the CBD-fusion proteins to a cellulose
 CC support. CBD-fusion proteins can be used to modify the chemical or
 CC physical properties of a cellulose or polysaccharide matrix column and to
 CC modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre.
 CC The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and
 CC has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDs from non-thermophilic organisms

SQ Sequence 562 AA;

Query Match 99.3%; Score 1974; DB 3; Length 562;
 Best Local Similarity 99.4%; Pred. No. 2.7e-178;
 Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRSMIDQIKSIGYNT 60
 Db 42 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRSMIDQIKSIGYNT 101
 Qy 61 IRLPYSDDILKPGCTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120
 Db 102 IRLPYSDDILKPGCTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 161
 Qy 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPVTVGFDLHNEPHDPACWCGDPSIDWRL 180
 Db 162 SGQSALWYTSVSEATWISDLQALAQRYKGNPVTVGFDLHNEPHDPACWCGDPSIDWRL 221
 Qy 181 AAERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGNGLOGAGQYPVVLNPNRLVYSAHDY 240
 Db 222 AAERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGNGLOGAGQYPVVLNPNRLVYSAHDY 281
 Qy 241 ATSVGPOTWFSDFPTFPNNPFGIWNKNGVLFNQNIAPIVNLGEGFTTLOSTTDQIWLKTLV 300
 Db 282 ATSVGPOTWFSDFPTFPNNPFGIWNKNGVLFNQNIAPIVNLGEGFTTLOSTTDQIWLKTLV 341
 Qy 301 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358
 Db 342 QYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 399

RESULT 6

AAB48787
 ID AAB48787 standard; protein; 521 AA.

AC AAB48787;
 DT 09-MAR-2001 (first entry)

XX Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.
 XX E1 endoglucanase; glycosyl hydrolase; soluble substrate;
 KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutain.

XX Acidothermus cellulolyticus.
 XX Synthetic.
 XX WO200070031-A1.
 XX 23-NOV-2000.
 XX 19-MAY-2000; 2000WO-US013971.
 XX 19-MAY-1999; 99US-0134925P.
 XX (MIDE) MIDWEST RES INST.

XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;
 XX Decker SR;
 XX WPI; 2001-061226/07.

XX Preparation of glycosyl hydrolase with an increased catalytic activity on
 PT insoluble substrate.
 XX Claim 17; Page 25-27; 30pp; English.
 XX The invention relates to a method for making glycosyl hydrolase mutants
 CC with increased catalytic activity with either insoluble or soluble
 CC cellulose substrates relative to the wild-type enzyme. The method for
 CC making a glycosyl hydrolase with increased insoluble substrate catalytic

XX PR 15-JUL-1994; 94US-00276213.
XX PA (MIDE) MIDWEST RES INST.
XX PI Thomas SR, Laymon RA, Himmel ME;
XX XX WPI; 1996-105843/11.
XX DR N-PSDB; AAT12337.
XX XX New isolated DNA encoding endo:glucanase - obtd from Acidothermus
XX PT cellulolyticus, used for prodn of the enzyme for use in cellulose
XX PT hydrolysis.
XX XX Claim 1; Page 22; 34pp; English.
XX XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful for
XX CC hydrolysing cellulosic biomass to sugars for simultaneous or subsequent
XX CC fermentation to ethanol. It shows optimal activity at 83 deg C. The amino
XX CC acid sequence of the mature enzyme was deduced from an E1 endoglucanase
XX CC gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides
XX CC were also identified (see also AAR89928-29). Cloning of this gene allows
XX CC large-scale, low-cost prodn. of recombinant E1 endoglucanase, using pref.
XX CC Saccharomyces, Zymomonas or E.coli hosts
XX XX Sequence 521 AA;
XX SQ
Query Match 99.5%; Score 1978; DB 2; Length 521;
Best Local Similarity 99.7%; Pred. No. 1e-178;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
QY 61 IRLPYSDILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120
DB 61 IRLPYSDILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120
QY 121 SQGSALWYTSVSEATWISDLQALQARYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 121 SQGSALWYTSVSEATWISDLQALQARYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240
QY 241 ATSGVPTWFSDDPTFPNNMPCGIWKNWGYLFNQNIAPVNLGEGTTLQSTTDQTLWLTIV 300
DB 241 ATSGVPTWFSDDPTFPNNMPCGIWKNWGYLFNQNIAPVNLGEGTTLQSTTDQTLWLTIV 300
QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGDTGGILXDDMQTVDTVKDGYLAPIKSIIDPV 358
DB 301 QYLRPTAQYAGDSFQWTFWSNPDSDGDTGGILXDDMQTVDTVKDGYLAPIKSIIDPV 358
RESULT 4
AAU79549 standard; protein; 562 AA.
XX XX AAU79549;
XX AC AAU79549;
XX XX 24-SEP-2002 (first entry)
XX DE A. cellulolyticus cellulase E1 beta-1,4-endoglucanase precursor.
XX XX Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;
XX KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;
XX KW industrial chemical; biodegradation; chloroacromatic;
XX KW environmental pollutant; E1 beta-1,4-endoglucanase; e1.
XX XX Acidothermus cellulolyticus.
XX OS
XX XX

XX FH Key Location/Qualifiers
XX FT Peptide 1..41
XX FT Protein /label= Leader_sequence
XX FT /label= Mature_cellulase
XX XX WO200234926-A2.
XX XX 02-MAY-2002.
XX XX 18-OCT-2001; 2001WO-US032538.
XX XX 20-OCT-2000; 2000US-0242408P.
XX XX (UNMS) UNIV MICHIGAN STATE.
XX XX Sticklen MB, Dale BE, Magbool S;
XX XX WPI; 2002-489947/52.
XX XX N-PSDB; ASK86729.
XX XX Producing transgenic plants which after harvest degrade lignin and
XX PT cellulose to fermentable sugars, by mating transgenic plant comprising
XX PT DNA encoding cellulase with transgenic plant comprising DNA encoding
XX PT ligninase.
XX XX Disclosure; Page 96-99; 126pp; English.
XX XX The invention discloses the production of a transgenic plant which
XX CC degrades lignocellulose when the plant is grown. It comprises the
XX CC production of the transgenic plant including cellulase and ligninase by
XX CC mating a transgenic plant, containing a DNA encoding a cellulase, and a
XX CC transgenic plant, containing a DNA encoding a ligninase, where both genes
XX CC are operably linked to a nucleotide sequence encoding a signal peptide
XX CC which targets the fusion protein to an organelle of the plant,
XX CC particularly chloroplasts. The method is useful for producing a
XX CC transgenic plant (e.g. maize) which degrades lignocellulose when the
XX CC plant is grown to produce a plant material. This material is useful for
XX CC converting lignocellulose, in a plant material, to fermentable sugars
XX CC which are then fermented to ethanol. The transgenic plants also provide a
XX CC plentiful and inexpensive source of fungal or bacterial cellulases and
XX CC ligninases which can be used in the production of ethanol. They can also
XX CC be used for pre-treating silage to increase the energy value of
XX CC lignocellulosic feeds for cows and other ruminant animals, pre-treating
XX CC lignocellulosic biomass for fermentative conversion to fuels and
XX CC industrial chemicals, and biodegradation of chloroacromatic environmental
XX CC pollutants. The protein sequence presented is the A. cellulolyticus
XX CC cellulase E1 beta-1,4-endoglucanase precursor
XX XX Sequence 562 AA;
XX SQ
Query Match 99.5%; Score 1978; DB 5; Length 562;
Best Local Similarity 99.7%; Pred. No. 1.1e-178;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60
DB 42 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 101
QY 61 IRLPYSDILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120
DB 102 IRLPYSDILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 161
QY 121 SQGSALWYTSVSEATWISDLQALQARYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180
DB 162 SQGSALWYTSVSEATWISDLQALQARYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 221
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240
DB 222 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 281
QY 241 ATSGVPTWFSDDPTFPNNMPCGIWKNWGYLFNQNIAPVNLGEGTTLQSTTDQTLWLTIV 300

CC (AA48788) has improved activity with insoluble substrates, and the W42R
 CC (AA48786) and Y62R (AA48787) mutants have improved activity with
 CC soluble substrates. The invention also encompasses DNA encoding these
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts
 CC for cellulose hydrolysis to produce sugars that can be fermented to
 CC produce fuels such as ethanol. The present sequence represents the
 CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant
 XX
 SQ Sequence 521 AA;

Query Match 100.0%; Score 1987; DB 4; Length 521;
 Best Local Similarity 100.0%; Pred. No. 1.4e-179;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
 DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
 QY 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120
 DB 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120
 QY 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
 DB 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
 QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVLNPNRLVYSAHDY 240
 DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVLNPNRLVYSAHDY 240
 QY 241 ATSVGPQTWFSPTFPNNMGGIWNKNWGLFNQNTAPVWLGEFGTTQSTTDTQWLKTLV 300
 DB 241 ATSVGPQTWFSPTFPNNMGGIWNKNWGLFNQNTAPVWLGEFGTTQSTTDTQWLKTLV 300
 QY 301 QYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358
 DB 301 QYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2
 AA39262
 ID AA39262 standard; protein; 358 AA.

XX AA39262;

XX 14-MAY-1998 (first entry)

XX A. cellulolyticus E1-CAT translated region.

XX E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;
 KW truncated; cellulose hydrolysis; biomass conversion.

XX Acidothermus cellulolyticus.

XX US5712142-A.

XX 27-JAN-1998.

XX 22-FEB-1996; 96US-00604913.

XX 26-SEP-1989; 89US-00412434.

XX 27-JAN-1992; 92US-00826089.

XX 21-SEP-1993; 93US-00125115.

XX 15-JUL-1994; 94US-00276213.

XX (MIDE) MIDWEST RES INST.

XX Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;

XX WPI; 1998-119985/11.

XX N-PSDB; AA09659.

XX DNA encoding truncated form of E1 endoglucanase from Acidothermus

PT only catalytic domain, is used for biomass conversion and has better heat
 PT stability than complete enzyme.

XX Claim 2; Fig 5; 19pp; English.

XX This sequence represents a novel Acidothermus cellulolyticus endoglucanase
 CC E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the
 CC enzyme, without its peptide linker or cellulose binding domain (CBD) and
 CC the truncated enzyme expressed by is used for hydrolysis of cellulose
 CC (biomass conversion). Compared with full-length E1, the truncated enzyme
 CC has better heat stability and higher temperature of maximum activity
 XX

SQ Sequence 358 AA;

Query Match 99.5%; Score 1978; DB 2; Length 358;
 Best Local Similarity 99.7%; Pred. No. 5.8e-179;
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
 DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT 60
 QY 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120
 DB 61 IRLPYSDDLKPGTWPNSINFYQMNQDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120
 QY 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
 DB 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180
 QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVLNPNRLVYSAHDY 240
 DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYWMGNLQAGQYPVVLNPNRLVYSAHDY 240
 QY 241 ATSVGPQTWFSPTFPNNMGGIWNKNWGLFNQNTAPVWLGEFGTTQSTTDTQWLKTLV 300
 DB 241 ATSVGPQTWFSPTFPNNMGGIWNKNWGLFNQNTAPVWLGEFGTTQSTTDTQWLKTLV 300
 QY 301 QYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358
 DB 301 QYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3

AA89927

ID AA89927 standard; protein; 521 AA.

XX AA89927;

XX 08-OCT-1996 (first entry)

XX A. cellulolyticus E1 endoglucanase.

XX E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers
 FT Domain 1.383
 FT Domain /label= Catalytic-domain
 FT Domain 364.417
 FT Domain /label= Linker
 FT Domain /note= "proline/serine/threonine-rich linker domain
 common to multi-domain microbial cellulases"

FT Domain 418.521
 FT Domain /label= Cellulose-binding_domain

XX WO9602551-A1.

XX 01-FEB-1996.

XX 14-JUL-1995; 95WO-US008868.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:04 ; Search time 47.6667 Seconds
(without alignments)
2122.070 Million cell updates/sec

Title: US-09-997-504A-10
Perfect score: 1987
Sequence: 1 AGGGVWHTSGREILDANNVP.....VDTKDGVLAPIKSSIFDPV 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesep 29Jan04: *
1: Genesep1980s: *
2: Genesep1990s: *
3: Genesep2000s: *
4: Genesep2001s: *
5: Genesep2002s: *
6: Genesep2003as: *
7: Genesep2003bs: *
8: Genesep2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1987	100.0	521	4	AAB48788 Acidother
2	1978	99.5	358	2	AAB39262 A. cellul
3	1978	99.5	521	2	AAR89927 A. cellul
4	1978	99.5	562	5	AAB79549 A. cellul
5	1974	99.3	562	3	AAB69508 Acidother
6	1969	99.1	521	4	AAB48787 Acidother
7	1964	98.8	521	4	AAB48786 Acidother
8	793.5	39.9	535	7	ADD24922 Xanthomon
9	763	38.4	628	2	AAB34999 Archaeabac
10	762	38.4	841	2	AAB34985 Archaeabac
11	762	38.3	553	5	ABG70759 T. mariti
12	755	38.0	430	3	AAB10344 P. horiko
13	755	38.0	458	3	AAB10345 P. horiko
14	682.5	34.3	514	4	AAB36264 Putative
15	621	31.3	616	2	AAY13494 Truncated
16	621	31.3	616	5	AAB16325 Active ce
17	621	31.3	1426	2	AAY13492 Truncated
18	621	31.3	1426	5	AAB16323 Active ce
19	513.5	25.8	425	4	ABB09060 Thermus c
20	276.5	13.9	582	7	ADD24920 Xanthomon
21	267	13.4	508	5	ABB2443 Herbicida
22	211.5	10.6	488	5	ABB3510 Herbicida
23	205	10.3	38	2	AAR89930 A. cellul
24	195	9.8	389	5	AAM49427 Penicilli
25	185	9.3	517	6	ABP99336 Orthosomy

26	183	9.2	551	5	ABB92442 Herbicida
27	173.5	8.7	782	6	ABP73022 Amino aci
28	171.5	8.6	375	6	ABP73019 Amino aci
29	160.5	8.1	722	6	ABU24069 Protein e
30	156.5	7.9	518	4	ABB06928 Micromono
31	153	7.7	420	6	ABP96833 Candida u
32	147.5	7.4	406	2	AAR08199 Neutral c
33	146	7.3	335	7	ADD24924 Ralstonia
34	143	7.2	400	4	AAE09785 Paenibaci
35	141.5	7.1	329	2	AAW35002 Thermotog
36	141	7.1	490	7	ADC58031 Endogluca
37	140	7.0	431	3	AAG42172 Arabidops
38	140	7.0	431	3	AAG27501 Arabidops
39	140	7.0	431	5	ABB93972 Herbicida
40	140	7.0	442	3	AAG42171 Arabidops
41	140	7.0	443	3	AAG27500 Arabidops
42	139	7.0	348	4	AAE09788 P. pabuli
43	139	7.0	476	3	AAAY54123 A. mannana
44	139	7.0	490	3	AAAY54122 Amino aci
45	138	6.9	335	4	AAE12786 Talaromyc

ALIGNMENTS

RESULT 1
AAB48788
ID AAB48788 standard; protein; 521 AA.

AC AAB48788;

DT 09-MAR-2001 (first entry)

DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.

XX E1 endoglucanase; glycosyl hydrolase; insoluble substrate;

KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutein.

XX Acidothermus cellulolyticus.

OS Synthetic.

XX WO200070031-A1.

XX PD 23-NOV-2000.

PF 19-MAY-2000; 2000WO-US013971.

PR 19-MAY-1999; 99US-0134925P.

XX (MIDE) MIDWEST RES INST.

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX Decker SR;

DR WPI; 2001-061226/07.

XX Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.

CC Claim 5; Page 27-29; 30pp; English.

CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic activity comprises replacing an active site-associated amino glycosyl-stabilising amino acid with an amino acid that does not strongly bind a disaccharide product in the active site. Conversely, the method for making a glycosyl hydrolase with increased soluble substrate catalytic activity comprises replacing a hydrophobic substrate-binding amino acid with a positively charged residue. The invention also discloses mutants of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788) produced according to the method of the invention. The Y245G mutant